GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 00:15:58 ; Search time 770.426 Seconds
(without alignments)
1744.014 Million cell updates/sec

Title: US-09-831-272-11
Perfect score: 31
Sequence: 1 tacaattcaaacattgttcaaacaaggaacc 31

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	31	100.0	31	6	AX025611	AX025611 Sequence
2	31	100.0	2683	8	PCPR2G	X55736 P.crispum P
3	22.8	73.5	257109	3	AC115577	AC115577 Dictyoste
c 4	22.6	72.9	224456	10	AC121833	AC121833 Mus muscu
5	22.2	71.6	101110	8	AP006141	AP006141 Lotus cor
c 6	22	71.0	112695	8	AC119418	AC119418 Medicago
c 7	22	71.0	130043	8	AC126019	AC126019 Medicago
8	22	71.0	168254	9	AC097502	AC097502 Homo sapi
9	22	71.0	176787	2	AC021374	AC021374 Homo sapi
10	22	71.0	217055	2	AC113221	AC113221 Rattus no
11	22	71.0	221944	2	AC107795	AC107795 Mus muscu
c 12	22	71.0	300680	2	AC129026	AC129026 Rattus no
13	21.6	69.7	1165	8	YSCMFAGB	M55016 S.italicus
c 14	21.6	69.7	37245	3	U41036	U41036 Caenorhabdi
c 15	21.6	69.7	41335	3	AC024750	AC024750 Caenorhab
16	21.6	69.7	103608	9	AL139161	AL139161 Human DNA
17	21.6	69.7	165146	2	AC011644	AC011644 Homo sapi
18	21.6	69.7	167163	2	AC013705	AC013705 Homo sapi
c 19	21.6	69.7	169337	2	AL356462	AL356462 Homo sapi
20	21.6	69.7	185182	9	AC093423	AC093423 Homo sapi
21	21.6	69.7	190599	2	AC146417	AC146417 Pan trogl
22	21.6	69.7	193131	9	AL390766	AL390766 Human DNA
c 23	21.6	69.7	209365	2	AC006719	AC006719 Caenorhab
c 24	21.6	69.7	248865	2	AC106672	AC106672 Rattus no
25	21.4	69.0	2000	6	AX596264	AX596264 Sequence
26	21.4	69.0	2000	6	AX819092	AX819092 Sequence
27	21.4	69.0	2000	6	AX830122	AX830122 Sequence
28	21.4	69.0	22131	8	SC9916	Z48952 S.cerevisia
c 29	21.4	69.0	56510	2	AC020215	AC020215 Drosophil
c 30	21.4	69.0	66170	2	AC010566	AC010566 Drosophil
c 31	21.4	69.0	122961	9	HSJ193N13	AL078600 Human DNA
c 32	21.4	69.0	136131	9	BS000063	BS000063 Pan trogl
33	21.4	69.0	150172	9	AC006285	AC006285 Homo sapi
34	21.4	69.0	166380	9	AC087237	AC087237 Homo sapi
35	21.4	69.0	166490	9	AL359438	AL359438 Human DNA
36	21.4	69.0	167682	2	AC119686	AC119686 Rattus no
c 37	21.4	69.0	174811	2	AC013259	AC013259 Homo sapi
38	21.4	69.0	180269	9	AC016025	AC016025 Homo sapi
39	21.4	69.0	182477	2	BX649274	BX649274 Danio rer
40	21.4	69.0	184657	3	AC010043	AC010043 Drosophil
c 41	21.4	69.0	194950	2	AC055115	AC055115 Homo sapi
42	21.4	69.0	208452	2	AC138581	AC138581 Pan trogl
c 43	21.4	69.0	216302	2	AC107116	AC107116 Rattus no
c 44	21.4	69.0	241384	2	AC094903	AC094903 Rattus no
c 45	21.4	69.0	274757	2	BX511000	BX511000 Danio rer

ALIGNMENTS

RESULT 1
AX025611
LOCUS AX025611 31 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 11 from Patent WO0029592.
ACCESSION AX025611
VERSION AX025611.1 GI:10187279
KEYWORDS .
SOURCE Petroselinum crispum (parsley)
ORGANISM Petroselinum crispum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Apiales; Apiaceae; Apioideae; apioid superclade; Apium clade; Petroselinum.
REFERENCE 1
AUTHORS Logemann,E., Somssich,I., Hahlbrock,K., Kirsch,C. and Rushton,P.

```
  TITLE     Chimeric promoters capable of mediating gene expression in plants
            upon pathogen infection and uses thereof
  JOURNAL   Patent: WO 0029592-A 11 25-MAY-2000;
            MAX PLANCK GESELLSCHAFT (DE) ; LOGEMANN ELKE (DE) ; SOMSSICH IMRE
            (DE) ; HAHLBROCK KLAUS (DE) ; KIRSCH CHRISTOPH (DE) ; RUSHTON PAUL
            (GB)
FEATURES             Location/Qualifiers
     source          1. .31
                     /organism="Petroselinum crispum"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:4043"
ORIGIN

  Query Match             100.0%;  Score 31;  DB 6;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 0.83;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGAACC 31
              ||||||||||||||||||||||||||||||
Db          1 TACAATTCAAACATTGTTCAAACAAGGAACC 31


RESULT 2
PCPR2G
LOCUS       PCPR2G                  2683 bp    DNA     linear   PLN 07-JUN-1991
DEFINITION  P.crispum PR2 gene for pathogenesis-related protein 2.
ACCESSION   X55736
VERSION     X55736.1  GI:20462
KEYWORDS    pathogenesis-related protein; PR2 gene.
SOURCE      Petroselinum crispum (parsley)
  ORGANISM  Petroselinum crispum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Apiales; Apiaceae; Apioideae; apioid
            superclade; Apium clade; Petroselinum.
REFERENCE   1  (bases 501 to 1800)
  AUTHORS   van de Locht,U., Meier,I., Hahlbrock,K. and Somssich,I.E.
  TITLE     A 125 bp promoter fragment is sufficient for strong
            elicitor-mediated gene activation in parsley
  JOURNAL   EMBO J. 9 (9), 2945-2950 (1990)
  MEDLINE   90361009
   PUBMED   2390976
REFERENCE   2  (bases 1 to 2683)
  AUTHORS   Somssich,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-1991) Somssich E., MAX-PLANCK-INSTITUTE F.
            ZUECHTUNGSFORSCHUNG, DEPT. OF BIOCHEMISTRY, CARL-VON-LINNE WEG 10,
            5000 KOELN 30, GERMANY
COMMENT     Evolutionarily conserved polypeptide; close association to disease
            resistance in plants.
            Origin of genomic library: Douglas et al. EMBO J. 6:189-195(1987).
FEATURES             Location/Qualifiers
     source          1. .2683
                     /organism="Petroselinum crispum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4043"
                     /clone="gPcPR2(Ps29)"
                     /cell_type="cultured parsley cells"
                     /clone_lib="lambda EMBL4 parsley genomic"
     promoter        1. .790
                     /evidence=experimental
     misc_feature    622. .739
                     /note="elicitor-responsive element"
                     /evidence=experimental
     CAAT_signal     700. .704
     TATA_signal     759. .766
     gene            791. .1785
                     /gene="PR2"
     mRNA            join(791. .1077,1273. .1785)
                     /gene="PR2"
                     /evidence=experimental
     prim_transcript 791. .>1754
                     /gene="PR2"
                     /evidence=experimental
     exon            791. .1077
                     /gene="PR2"
                     /number=1
     CDS             join(894. .1077,1273. .1565)
                     /gene="PR2"
                     /codon_start=1
                     /evidence=experimental
                     /product="pathogenesis-related protein 2"
                     /protein_id="CAA39268.1"
                     /db_xref="GI:20463"
                     /db_xref="GOA:P27538"
                     /db_xref="SWISS-PROT:P27538"
                     /translation="MGAVTTDVEVASSVPAQTIYKGFLLDMDNIIPKVLPQAIKSIEI
                     ISGDGGAGTIKKVTLGEVSQFTVVKQRIDEIDAEALKYSYSIIEGDLLLGIIESITSK
                     FTVVPTDGGCIVKNTTIYTPIGDAVIPEENVKEATEQSGMVFKAIEAYLLANPGAY"
     intron          1078. .1272
                     /gene="PR2"
                     /number=1
     exon            1273. .1785
                     /gene="PR2"
                     /number=2
     polyA_signal    1749. .1754
                     /gene="PR2"
ORIGIN

  Query Match             100.0%;  Score 31;  DB 8;  Length 2683;
  Best Local Similarity   100.0%;  Pred. No. 0.25;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGAACC 31
              ||||||||||||||||||||||||||||||
Db        715 TACAATTCAAACATTGTTCAAACAAGGAACC 745
```

GI 20462

```
RESULT 3
AC115577
LOCUS       AC115577               257109 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map 4657875-4914984 strain
            AX4, complete sequence.
ACCESSION   AC115577
VERSION     AC115577.2  GI:28829140
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 257109)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
  MEDLINE   22092622
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 257109)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 257109)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 257109)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Mar 4, 2003 this sequence version replaced gi:19569889.
            CDS predictions from GeneID do not necessarily reflect true genes.
```

```
        Further Information is available from IMB Jena, Department of
        Genome Analysis
        (http://genome.imb-jena.de/dictyostelium/)
        and the Univerity Colonge, Institute for Biochemistry I
        (http://www.uni-koeln.de/dictyostelium/project.shtml
        Funding
        Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES        Location/Qualifiers
     source     1. .257109
                /organism="Dictyostelium discoideum"
                /mol_type="genomic DNA"
                /strain="AX4"
                /db_xref="taxon:44689"
                /chromosome="2"
                /map="4657875-4914984"
     CDS        join(1913. .1925,2047. .2279)
                /note="GeneID exon scores (in order of location ranges):
                2.72, 10.68 - GSCJ_ID dd_02172"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO51694.1"
                /db_xref="GI:28829141"
                /translation="MTIINNISTFGSKNNMANATQKIDFQNTTNQLNTSNNISNSLQS
                SAKINTTCDNSCSQNDIIIGGYKLPKSSELPYDCTIC"
     CDS        complement(2661. .3743)
                /note="GeneID exon scores (in order of location ranges):
                74.49 - GSCJ_ID dd_02173"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAL92208.2"
                /db_xref="GI:28829142"
                /translation="MKNIKIIIILISLIFLISISNSEIFGRSEQCPLISLLKIGEDCQ
                SSSDCEYPNICKNNKCQEIVQVGDPCESDEECGLFFNNIGKCISGTCDINLKNGDDCG
                EGHSGTCGEGSVCKFYKCVLNNGNCTSDSQCPFNHYCEENSNTCTPVVLGNQCDRDSM
                CPITHICQKNECVAKYSRETATPCGGNPELCDVFNGDECDKVHKICRRNDKYGSECYV
                DGDCGTGVCICNNDGHKQCVGSNTTLVHDGCLQAINSFVKCMKDKRCSSVSPLTCKDC
                YPIWECYRYNCFASEHFDQRKSAYYKSLKCDTDVGSLVNPDAANPNYNGGSNRKTNQK
                RSNSNSNYLISLSLTLTLILILLILF"
     CDS        complement(join(4345. .7339,7520. .7677))
                /note="GeneID exon scores (in order of location ranges):
                282.51, 3.60 - GSCJ_ID dd_02175"
                /codon_start=1
                /product="similar to Oryza sativa (japonica
                cultivar-group). OJ1513_F02.22 protein"
                /protein_id="AAO51695.1"
                /db_xref="GI:28829143"
                /translation="MINNSPPINIKSENIYQYQIQSIEYYNNEKYVECMNILNKLIEL
                NTNDIIALCNRATCNYKLELYKHCIKDCEVILCKEPNNKIALLRKAKALKQLSNYDIE
                LVNSLRKKANNSWGYLSVTIDDNDDGDDNNINSNINNINNNNNNNNNKNISISSPKTT
                TTSSIFISTSPPPYLSQQSQQQFHTMTIREQQQQQQQFLQEQQMLIYQEQQQQFLQEQ
                QEMIDDLLSSPPLSLSEIHQKQHIQGDGNNIGEEEEDEEEEGEEGGIIVKSDSMEELN
                KLLEQLSTENIATLISQGNTFVNAGQYEEAIELFTMIIDNNPLVPSLYLGRGTSNAFL
                GQLNEAINDFSRAIELDNTSSDAYKRRGQSKVAKSMEQEALEDFNQAVAFDKEDDYDI
                YYNRGLLHYQMRNYERALKDFKKVTSIEPSHKLAWNRIGLCLNVNGYPMEAYQAFTEA
                LRIDPYFEASHTNIGQCWKDLGKFDPSYQSFTKALEICPNYSNALHLRGLLFFNSGRH
                LDAIKDWNLFLKLYGSGGANSSGSNIKTGNHYSGGGGGGGGGEELSSIIDVRQFRGIV
                NQSIGAFRKAVADYDQILQTNPLHNCFYQREIALWTHHHLNTPLSTFNIDRFINCYLK
                TFQCQRIPNTSLTCNSISLGFQLQSSFNQSVADVEICHQLNNEQLSLFKEAKRIGEML
                QYNCDGFLTNSRQKLQCGLSVIEMAQTLKKLWLSGNKKSSVFQLDGKSSSVSTKKHTF
                SWRDLYDIGVKWRQISEPNDPVWWGDLLSPEQFKEGFGSHTPIITGQCKVVRYYPMFE
                KSFEIMKLLLPIQHPSLISNQTILDQLESATTCKDLYKLIKKDFFVVTPCYSRSNENK
                VMEGTRLTLQYIHPEGYEFAIRTPSTPYRWEQYSQELDIIWDQLIDKVFIYQNIINSK
                EDLIDNSSSSSSSSPSPSPSNQESILLDEISDLILCLSFYWYNFMPLSRGSAALGFTT
                LLGLFLSLDIAITSSIPKNQQPDWEAILRPTTTSFISVIKPWLYPARSKFNLQTTSLV
                SSQFNTIKKMIQVLNNSNQY"
     CDS        complement(join(8323. .9266,9382. .9403))
                /note="GeneID exon scores (in order of location ranges):
                87.76, 4.06 - GSCJ_ID dd_02178"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO51696.1"
                /db_xref="GI:28829144"
                /translation="MSTMKAAVLYEIGTSPKYDEFPEPKEHGDFVKVSMIASTIKQLD
                RGRVSGRHYMKYTKFPTAVGIDGVGRLEDGQLVYCFGDGMIAEKSLVEKGKYAPLSND
                VDPYIAAALPNTLLGSDMALLFRAEIKKGDVVLVNGATGVTGSVAVQIAKIRGASKVI
                AAGRNEEALKKLQADYGVDEILVLNSEEGVAAQMKQILDKQPIDIVLDYLWGQPTEMI
                IKSIVAQKQYHVTKIVSVGEMAGSSINLPSAALRSTAISILGSGLGSFPATQMHTYYK
                EHLNELLLLAKQGKLKVDIKIIDLKDIEKEWLVNEKSVRTVVKIN"
     CDS        join(10352. .10495,10565. .12829)
                /note="GeneID exon scores (in order of location ranges):
                11.83, 249.82 - GSCJ_ID dd_02181"
                /codon_start=1
                /product="similar to Giardia lamblia (Giardia
                intestinalis). Arginine deiminase (EC 3.5.3.6)"
                /protein_id="AAO51697.1"
                /db_xref="GI:28829145"
                /translation="MKQSISNYQLANMASPFPTKGPLYNLSGKYKDLSFNQVHENDEA
                RIVILNEPSLPQWMGSIHPNGSLYEEPLDLSKAKEEHKLFRELLEKEGCTVFTVREIL
                TGSYDQTKSEDIQHRVKLEDFAFNSMHYELDSQQNVQELSTEDKLLISDQYKKKCIEC
                MSNEQLVEVILTRPTIRLRKSERDTELLSTEYSFKPLVNLIFQRDQQITTAAGIVLAS
                LSSPIRAPEVQVMKLCFEILGLPIVGEIPEPGKLEGGDFYPVGRDLSLIGVGLRSNFY
                SVQYMMENDLFGTTRVAVVKDYFDQNQQRMHLDTVCNIIDERVMLILEDICGEDSPIR
                RLVDEYTRDDEGNYHLTRNDVEYSQYLKEEGFDLIYASNQNQKDYGCNGLNIGNGKFI
                VVDRATAKTIARSTKTNSQLLFVDFRTVTKMYGSVHCCSQVVSRFDRPRPGGSLQHSP
                SSDKLSPQIIRRNSRDSSSGSGSDQDEDGSDKFLMISPTDFVKDRIGMNASLIRQQRQ
                LILQDYSKLHKLLVDAGVRVHLFCHEPYHRTDKAVFVADWFSTHDRDEIKSTEKNTLV
                LYPLSNQSRRRERRADILKFGFLPKYTRVLDFTPCEQGKLEYNETTIEQFIKDSSSLD
                ESSSPSINGDDGYFIDFSEIVLDREHRLAYCSTKNDRQSKFIDLWASALGYKVIKINS
                KEAANKFIFIGSKFALFCPNALSSKEQSEEVIQALQTTSAKKVIEITEEQMNNSCAKI
                IQVKGNQKTHLLMSDKVMNSFTSNQLTELKSYVDQLLHADMSNLENMGAHNDGMISML
                GRLV"
     CDS        join(13329. .14096,14186. .14425)
                /note="GeneID exon scores (in order of location ranges):
                63.00, 17.23 - GSCJ_ID dd_02185"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO51698.1"
                /db_xref="GI:28829146"
                /translation="MTEDYIRLFSQTAIADLDKNTMVSIKKSEDVEKLLSVLYENSLS
                SLPVIDGEGKRVIGFVDTNDVLSLLAKLKLSSDSSDDDIRLYSIGFLNSNVNNIMDEK
                KKDQFVVVLEEQSLFEVLKLYANGVHRLALLTVFSDIENIVSQSNVIKFLNNNLSVLG
                QLGDTTIGSSLIKPFLPSKESLITTKSSTLTIDSFKIMNTHKCSAIPVLEDSTNKIIG
                TLSVNDLYGINQSTIKLLLNPTGEFINLDNSKIKSNKNKPNHQIVLKLNNTFKEAIQI
                ISNNKVHRIWIVDDNNIPISLISLTDICKIIVEAPYLLEDQQTNNNQRNSDNSTTNQN
                K"
     CDS        complement(join(14882. .15047,15132. .15385))
                /note="GeneID exon scores (in order of location ranges):
                9.86, 24.37 - GSCJ_ID dd_02189"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO51699.1"
                /db_xref="GI:28829147"
                /translation="MKEFDQLLSVESDAYISSKEQGIDDGKRLGYVEGYQLGFEKGIE
                LGQEIGYYQSCVTVWNHLVSINNNNNNNNNNNKNNLKFSVRGIQNLEKLTKLLEDYHL
                DFNDENIMNTLSEIRLKFKLTSVQLGLQTKENDELSF"
     CDS        join(15575. .16107,16207. .16540,16648. .17625)
                /note="GeneID exon scores (in order of location ranges):
                43.82, 14.38, 70.32 - GSCJ_ID dd_02192"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO51700.1"
                /db_xref="GI:28829148"
                /translation="MESIKSNNKGTCWFDWDEWKNIYNKLYSTDENEQRKAIKRLSAW
                RSRNKLPISIEVTGYFVELNLNRNCRSQNELELGLSMAISRMVNGFVDSTRDILKLNK
                INMFKKANEINLPTMFVDIRHESSHGKLPHLQLLKPISITALDWLNDYYWKPQLTHLH
                SHFQKIRSFLNIYKIQSEEDNGFIRYIEESMELFNDISSDILGNIMIPILIDEDYLIS
                NEFLPQTNSYAHIPSQLKKIYQPLLQFLHSKYPHFLIIILYLLIDRICKFKFLPPQSK
                KQKSTPIPLNDNQIIESKISLSMTHWTLYFLQEFNITVKKQEGLVNYKIDLPYRHILEK
                CIENIHSKYCQKILNVAMKKNLDGSKRVYTTATNKKKNKLNSATISNLDSTDLDLDPD
                EKINDSGEPQINYSVDQVLSNLNSNNNSSDIKNLSTSTSTSKSTKWKLQNQWISSPIG
                VLPPPNDQNSSTFDLLDDLDDNLDLGNFVEVSSVDYTTKDPSLHNSQSLFSKIKNIYP
                NLNKNSIIRKEFVNEIKNEKENEKEKENKNEKENKNENENEKNKDENENNKDENEKNE
                EDDDVIEIPQPILKKQKTSTTSSVPKKNQQKGKTKGKSNALDDFLFWK"
     CDS        complement(17727. .18857)
                /note="GeneID exon scores (in order of location ranges):
                64.81 - GSCJ_ID dd_02171"
```

```
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO51701.1"
                     /db_xref="GI:28829149"
                     /translation="MNNFPWEDFKAAARTGKTENTDKYSQMLILTLSNYLIEPPLVPI
                     NESCKPEPLPIINNSVCGKYPRVFSGKRDKPVKIGHMVQIGFDIDVLEVHLNELYDVI
                     DHFFILESTVTHYHKMKKPLIWEHVKFQDRFIKFQDKVVHLILDDTDEENGKGLFDAE
                     SYQETRRWQKFLYWNERTNQYGDDDIIGFGDTDEISARINLHYLKNCQFRKSVDIVDI
                     GIWFPYGPINQVFKPFYSVPNNPYTLGDPTFYTLKRAKSTTKAPSRNRGTSGHYMLGG
                     MHMTHYGYLPFQMVKYLSCTECGIKNETHVKNFSDDIQNGNIQKLELRLLETKSDHAY
                     KIQDISTMDDFFKKEVAILPWFYNCNRNRYPVWERKNDPRLF"
     CDS             complement(join(19753. .20363,20450. .21625,21828. .21885))
                     /note="GeneID exon scores (in order of location ranges):
                     42.45, 78.19, -3.29 - GSCJ_ID dd_03249"
                     /codon_start=1
                     /product="similar to Plasmodium falciparum. Hypothetical
                     protein"
                     /protein_id="AAO51702.1"

  Query Match             73.5%;  Score 22.8;  DB 3;  Length 257109;
  Best Local Similarity   92.3%;  Pred. No. 75;
  Matches   24;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAG 26
              |||||||||||||||||| || ||||
Db     228995 TACAATTCAAACATTGTTTAATCAAG 229020


RESULT 4
AC121833/c
LOCUS       AC121833               224456 bp    DNA     linear   ROD 27-NOV-2003
DEFINITION  Mus musculus BAC clone RP24-69C16 from chromosome 3, complete
            sequence.
ACCESSION   AC121833
VERSION     AC121833.3  GI:30425604
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 224456)
  AUTHORS   Abbott,S., Haakenson,W. and Schatzkamer,K.
  TITLE     The sequence of Mus musculus BAC clone RP24-69C16
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 224456)
  AUTHORS   Wilson,R.
  TITLE     Sequencing of Mus musculus
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 224456)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 224456)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 224456)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   6  (bases 1 to 224456)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On May 8, 2003 this sequence version replaced gi:22475741.
            ---------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            ---------- Summary Statistics
                Center project name: M_BB0069C16
            ----------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-24 BAC Library has been constructed by Pieter de Jong and
            coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
            and/or brain genomic DNA. The clone and detailed information can be
            obtained from Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone. This clone is
            overlapped by AC123057.
FEATURES                Location/Qualifiers
     source             1. .224456
                        /organism="Mus musculus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10090"
                        /chromosome="3"
                        /map="3"
                        /clone="RP24-69C16"
                        /clone_lib="RPCI-24"
     repeat_region      5. .193
                        /rpt_family="L1"
     repeat_region      406. .462
                        /rpt_family="ERVK"
     repeat_region      468. .659
                        /rpt_family="B2"
     repeat_region      668. .792
                        /rpt_family="Alu"
     repeat_region      883. .1005
                        /rpt_family="B4"
     repeat_region      1568. .2199
                        /rpt_family="L1"
     repeat_region      2281. .2427
                        /rpt_family="MER1_type"
     repeat_region      2431. .2700
                        /rpt_family="L1"
     repeat_region      2797. .2891
                        /rpt_family="B4"
     repeat_region      2906. .2961
                        /rpt_family="ID"
     repeat_region      2962. .3019
                        /rpt_family="B4"
     repeat_region      3081. .3418
                        /rpt_family="MaLR"
     repeat_region      3476. .3538
                        /rpt_family="RMER19A"
     repeat_region      3526. .3824
                        /rpt_family="RMER19B"
     repeat_region      3859. .5045
                        /rpt_family="ERVK"
     repeat_region      5252. .5431
```

```
                     /rpt_family="B2"
     repeat_region   5523. .5669
                     /rpt_family="Alu"
     repeat_region   5674. .5857
                     /rpt_family="ERVK"
     repeat_region   6045. .6188
                     /rpt_family="Alu"
     repeat_region   6325. .6531
                     /rpt_family="B2"
     repeat_region   6582. .6648
                     /rpt_family="Alu"
     repeat_region   7934. .8267
                     /rpt_family="ERVK"
     repeat_region   8361. .8539
                     /rpt_family="B2"
     repeat_region   8713. .8764
                     /rpt_family="ID"
     repeat_region   8890. .9036
                     /rpt_family="MER1_type"
     repeat_region   9041. .9228
                     /rpt_family="B2"
     repeat_region   9239. .9367
                     /rpt_family="B4"
     repeat_region   9443. .9584
                     /rpt_family="Alu"
     repeat_region   9586. .9715
                     /rpt_family="L1"
     repeat_region   9718. .10040
                     /rpt_family="L1"
     repeat_region   10042. .10125
                     /rpt_family="RMER6A"
     repeat_region   10126. .10271
                     /rpt_family="Alu"
     repeat_region   10272. .10410
                     /rpt_family="RMER6A"
     repeat_region   10382. .10480
                     /rpt_family="RMER6A"
     repeat_region   10481. .10655
                     /rpt_family="B2"
     repeat_region   10656. .10758
                     /rpt_family="RMER6A"
     repeat_region   10937. .11402
                     /rpt_family="ERVL"
     repeat_region   11769. .11941
                     /rpt_family="L1"
     repeat_region   12103. .12296
                     /rpt_family="L1"
     repeat_region   12321. .12509
                     /rpt_family="L1"
     repeat_region   12571. .12708
                     /rpt_family="Alu"
     repeat_region   12977. .13082
                     /rpt_family="MaLR"
     repeat_region   13083. .13234
                     /rpt_family="Alu"
     repeat_region   13235. .13267
                     /rpt_family="MaLR"
     repeat_region   13405. .13576
                     /rpt_family="L2"
     repeat_region   13746. .13840
                     /rpt_family="Alu"
     misc_feature    14289. .15623
                     /note="CpG_island (%GC=73.8, o/e=0.88, #CpGs=177)"
     repeat_region   16567. .16757
                     /rpt_family="B2"
     tRNA            complement(16681. .16752)
                     /product="tRNA-Ser"
                     /note="Likely pseudogene (HMM Sc=32.94 / Sec struct
                     Sc=-11.27)"
     repeat_region   16886. .16988
                     /rpt_family="B4"
     repeat_region   17335. .17444
                     /rpt_family="MIR"
     repeat_region   17661. .17811
                     /rpt_family="Alu"
     repeat_region   17832. .17932
                     /rpt_family="Alu"
     repeat_region   18209. .18355
                     /rpt_family="Alu"
     repeat_region   18377. .18558
                     /rpt_family="B2"
     repeat_region   18852. .18931
                     /rpt_family="MER1_type"
     repeat_region   18956. .19071

  Query Match            72.9%;  Score 22.6;  DB 10;  Length 224456;
  Best Local Similarity  86.2%;  Pred. No. 93;
  Matches   25;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAAC 30
              ||  ||||||| ||||||||| |||||||
Db     113314 ACTGTTCAAACCTTGTTCAAAAAAGGAAC 113286


RESULT 5
AP006141
LOCUS       AP006141              101110 bp    DNA     linear   PLN 22-JUL-2003
DEFINITION  Lotus corniculatus var. japonicus genomic DNA, chromosome 4,
            clone:LjT01N13, TM0247, complete sequence.
ACCESSION   AP006141
VERSION     AP006141.1  GI:29122778
KEYWORDS    HTG.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Kato,T., Sato,S., Nakamura,Y. and Tabata,S.
  TITLE     Structural analysis of a Lotus japonicus genome. III. Sequence
            features and mapping of sixty-two TAC clones which cover the 6.7 Mb
            regions of the genome
  JOURNAL   DNA Res. 10 (1), 27-33 (2003)
  MEDLINE   22579290
   PUBMED   12693552
REFERENCE   2  (bases 1 to 101110)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2002) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1. .101110
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="4"
                     /clone="LjT01N13"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM0247~synonym: Lotus japonicus"
ORIGIN

  Query Match            71.6%;  Score 22.2;  DB 8;  Length 101110;
  Best Local Similarity  88.9%;  Pred. No. 1.6e+02;
  Matches   24;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 CAATTCAAACATTGTTCAAACAAGGAA 29
              ||||||| | ||||||||| |||||||
Db      93667 CAATTCACAAATTGTTCAAGCAAGGAA 93693


RESULT 6
```

```
AC119418/c
LOCUS       AC119418              112695 bp    DNA     linear   PLN 29-APR-2003
DEFINITION  Medicago truncatula clone mth1-23116, complete sequence.
ACCESSION   AC119418
VERSION     AC119418.5  GI:30172592
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 112695)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth1-23116
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 112695)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 112695)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-AUG-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 112695)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5  (bases 1 to 112695)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   6  (bases 1 to 112695)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Apr 29, 2003 this sequence version replaced gi:24371410.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
FEATURES             Location/Qualifiers
     source          1. .112695
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /clone="mth1-23116"
                     /clone_lib="Medicago truncatula BAC library H1"
ORIGIN

  Query Match             71.0%;  Score 22;  DB 8;  Length 112695;
  Best Local Similarity   83.3%;  Pred. No. 1.8e+02;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGAAC 30
              | |||||| || |||| | ||||||||||||
Db     106456 TTCAATTCTAAAATTGGTGAAACAAGGAAC 106427


RESULT 7
AC126019/c
LOCUS       AC126019              130043 bp    DNA     linear   PLN 22-APR-2003
DEFINITION  Medicago truncatula clone mth2-22p22, complete sequence.
ACCESSION   AC126019
VERSION     AC126019.15  GI:30039760
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 130043)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2-22p22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 130043)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 130043)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 130043)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-APR-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Apr 22, 2003 this sequence version replaced gi:29893702.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
FEATURES             Location/Qualifiers
     source          1. .130043
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /clone="mth2-22p22"
                     /clone_lib="Medicago truncatula BAC library H2"
ORIGIN

  Query Match             71.0%;  Score 22;  DB 8;  Length 130043;
  Best Local Similarity   83.3%;  Pred. No. 1.8e+02;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGAAC 30
              | |||||| || |||| | ||||||||||||
Db      51160 TTCAATTCTAAAATTGGTGAAACAAGGAAC 51131


RESULT 8
AC097502
LOCUS       AC097502              168254 bp    DNA     linear   PRI 20-APR-2002
DEFINITION  Homo sapiens BAC clone RP11-340B18 from 4, complete sequence.
ACCESSION   AC097502 AC021465
VERSION     AC097502.3  GI:20143836
KEYWORDS    HTG.
```

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168254)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 168254)
  AUTHORS   Boatright,E., Haglund,K. and Elliott,G.
  TITLE     The sequence of Homo sapiens BAC clone RP11-340B18
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 168254)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-OCT-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 168254)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 168254)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Apr 12, 2002 this sequence version replaced gi:17647070.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0340B18
                Drafting Center: WIBR
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-511B7; the clone sequenced
            to the right is RP11-371E22, 2000 bp overlap.  Actual start of this
            clone is at base position 1 of RP11-340B18; actual end is at base
            position 11859 of RP11-371E22.

            The sequence of AC021465 has been incorporated into AC097502.
FEATURES             Location/Qualifiers
     source          1. .168254
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-340B18"
                     /clone_lib="RPCI-11"
     repeat_region   70. .551
                     /rpt_family="L2"
     repeat_region   599. .841
                     /rpt_family="MIR"
     repeat_region   858. .925
                     /rpt_family="L2"
     repeat_region   1282. .1588
                     /rpt_family="ERVL"
     repeat_region   1646. .1993
                     /rpt_family="L2"
     repeat_region   1989. .2543
                     /rpt_family="L2"
     repeat_region   2561. .2986
                     /rpt_family="MaLR"
     repeat_region   3017. .3349
                     /rpt_family="L2"
     repeat_region   3350. .3618
                     /rpt_family="ERVL"
     repeat_region   3686. .3817
                     /rpt_family="ERVL"
     repeat_region   4099. .4446
                     /rpt_family="MaLR"
     repeat_region   4533. .4703
                     /rpt_family="L2"
     repeat_region   4819. .5670
                     /rpt_family="ERVL"
     repeat_region   6274. .6451
                     /rpt_family="ERVL"
     repeat_region   6603. .6767
                     /rpt_family="ERVL"
     repeat_region   6768. .7164
                     /rpt_family="L2"
     repeat_region   7171. .7194
                     /rpt_family="(TA)n"
     repeat_region   7195. .7487
                     /rpt_family="Alu"
     repeat_region   7493. .7630
                     /rpt_family="Alu"
     repeat_region   7632. .7798
                     /rpt_family="L2"
     repeat_region   7801. .7921
                     /rpt_family="L2"
     repeat_region   10316. .10646
                     /rpt_family="MaLR"
     repeat_region   11599. .12022
                     /rpt_family="ERVL"
     repeat_region   12421. .12507
                     /rpt_family="ERVL"
     repeat_region   12559. .12627
                     /rpt_family="L2"
     repeat_region   12767. .12864
                     /rpt_family="MIR"
     repeat_region   12865. .13162
                     /rpt_family="Alu"
     repeat_region   13163. .13261
                     /rpt_family="MIR"
     repeat_region   13352. .13600
                     /rpt_family="MIR"
     repeat_region   14433. .14872
```

```
                     /rpt_family="L1"
     repeat_region   14873. .15166
                     /rpt_family="Alu"
     repeat_region   15167. .15208
                     /rpt_family="L1"
     repeat_region   15548. .15846
                     /rpt_family="L2"
     repeat_region   16186. .16368
                     /rpt_family="MER1_type"
     repeat_region   16369. .16681
                     /rpt_family="Alu"
     repeat_region   16682. .16817
                     /rpt_family="MER1_type"
     repeat_region   17980. .18009
                     /rpt_family="AT_rich"
     repeat_region   19802. .19927
                     /rpt_family="L2"
     repeat_region   21923. .22157
                     /rpt_family="L1"
     repeat_region   22158. .22444
                     /rpt_family="Alu"
     repeat_region   22445. .22503
                     /rpt_family="L1"
     repeat_region   22928. .22958
                     /rpt_family="AT_rich"
     repeat_region   23038. .23062
                     /rpt_family="AT_rich"
     repeat_region   23318. .23612
                     /rpt_family="Alu"
     repeat_region   23901. .24158
                     /rpt_family="L1"
     repeat_region   24159. .24427
                     /rpt_family="Alu"
     repeat_region   24428. .26531
                     /rpt_family="L1"
     repeat_region   26548. .26585
                     /rpt_family="AT_rich"
     repeat_region   26710. .27498
                     /rpt_family="L1"
     repeat_region   27501. .27915
                     /rpt_family="MaLR"
     repeat_region   28335. .28518
                     /rpt_family="L1"
     repeat_region   28625. .29170
                     /rpt_family="L1"
     repeat_region   29171. .29659
                     /rpt_family="MaLR"
     repeat_region   29660. .29767
                     /rpt_family="L1"

  Query Match             71.0%;  Score 22;  DB 9;  Length 168254;
  Best Local Similarity   83.3%;  Pred. No. 1.7e+02;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAACC 31
              ||| ||||||||   |||||||||||||  |
Db       9668 ACAGTTCAAACAAGTTTCAAACAAGGAAAC 9697


RESULT 9
AC021374
LOCUS       AC021374              176787 bp    DNA     linear   HTG 16-MAR-2000
DEFINITION  Homo sapiens clone RP11-25N12, WORKING DRAFT SEQUENCE, 15 unordered
            pieces.
ACCESSION   AC021374
VERSION     AC021374.4  GI:7249071
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176787)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-25N12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176787)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 16, 2000 this sequence version replaced gi:6939810.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4604
                Center clone name: 25_N_12
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 165864 bases at least Q40
                Consensus quality: 170217 bases at least Q30
                Consensus quality: 172405 bases at least Q20
                Insert size: 176000; agarose-fp
                Insert size: 175387; sum-of-contigs
                Quality coverage: 4.8 in Q20 bases; agarose-fp
                Quality coverage: 4.9 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1339: contig of 1339 bp in length
            *     1340     1439: gap of 100 bp
            *     1440     4249: contig of 2810 bp in length
            *     4250     4349: gap of 100 bp
            *     4350     7864: contig of 3515 bp in length
            *     7865     7964: gap of 100 bp
            *     7965    11523: contig of 3559 bp in length
            *    11524    11623: gap of 100 bp
            *    11624    16505: contig of 4882 bp in length
            *    16506    16605: gap of 100 bp
            *    16606    22103: contig of 5498 bp in length
            *    22104    22203: gap of 100 bp
            *    22204    27892: contig of 5689 bp in length
            *    27893    27992: gap of 100 bp
            *    27993    34300: contig of 6308 bp in length
            *    34301    34400: gap of 100 bp
            *    34401    43494: contig of 9094 bp in length
            *    43495    43594: gap of 100 bp
```

```
            *    43595    55836: contig of 12242 bp in length
            *    55837    55936: gap of 100 bp
            *    55937    72414: contig of 16478 bp in length
            *    72415    72514: gap of 100 bp
            *    72515    95313: contig of 22799 bp in length
            *    95314    95413: gap of 100 bp
            *    95414   117557: contig of 22144 bp in length
            *   117558   117657: gap of 100 bp
            *   117658   143798: contig of 26141 bp in length
            *   143799   143898: gap of 100 bp
            *   143899   176787: contig of 32889 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176787
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-25N12"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1339
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1440. .4249
                     /note="assembly_fragment"
     misc_feature    4350. .7864
                     /note="assembly_fragment"
     misc_feature    7965. .11523
                     /note="assembly_fragment"
     misc_feature    11624. .16505
                     /note="assembly_fragment"
     misc_feature    16606. .22103
                     /note="assembly_fragment"
     misc_feature    22204. .27892
                     /note="assembly_fragment"
     misc_feature    27993. .34300
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    34401. .43494
                     /note="assembly_fragment"
     misc_feature    43595. .55836
                     /note="assembly_fragment"
     misc_feature    55937. .72414
                     /note="assembly_fragment"
     misc_feature    72515. .95313
                     /note="assembly_fragment"
     misc_feature    95414. .117557
                     /note="assembly_fragment"
     misc_feature    117658. .143798
                     /note="assembly_fragment"
     misc_feature    143899. .176787
                     /note="assembly_fragment"
ORIGIN

  Query Match             71.0%;  Score 22;  DB 2;  Length 176787;
  Best Local Similarity   83.3%;  Pred. No. 1.6e+02;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAACC 31
              ||| ||||||||   |||||||||||| |
Db     103158 ACAGTTCAAACAAGTTTCAAACAAGGAAAC 103187


RESULT 10
AC113221
LOCUS       AC113221               217055 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-240B21, WORKING DRAFT SEQUENCE.
ACCESSION   AC113221
VERSION     AC113221.5  GI:25072721
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 217055)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 217055)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 217055)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23816990.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
```

```
          by sized gaps filled with Ns to the estimated size. The sequence
          may extend beyond the ends of the clone and there may be sequence
          contigs within a contig-scaffold that consist entirely of whole
          genome shotgun sequence reads. Both end sequences and whole genome
          shotgun sequence only contigs will be indicated in the feature
          table.
          -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
              Center project name: GSGU
              Center clone name: CH230-240B21
          -------------- Summary Statistics
              Assembly program: Phrap; version 0.990329
              Consensus quality: 210001 bases at least Q40
              Consensus quality: 211814 bases at least Q30
              Consensus quality: 213376 bases at least Q20
              Estimated insert size: 217709; sum-of-contigs estimation
              Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 1 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *        1   217055: contig of 217055 bp in length.
FEATURES             Location/Qualifiers
     source          1. .217055
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-240B21"
     misc_feature    1. .1949
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    3293. .4127
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ106860"
ORIGIN

  Query Match             71.0%;  Score 22;  DB 2;  Length 217055;
  Best Local Similarity   83.3%;  Pred. No. 1.6e+02;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAACC 31
              |||||||||| |||||||||  || | |||
Db     164275 ACAATTCAAATATTGTTCAACAAAAGGACC 164304


RESULT 11
AC107795
LOCUS       AC107795               221944 bp    DNA     linear   HTG 19-DEC-2003
DEFINITION  Mus musculus chromosome 1 clone RP23-433P9 map 1, *** SEQUENCING IN
            PROGRESS ***, 14 unordered pieces.
ACCESSION   AC107795
VERSION     AC107795.7  GI:40217727
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 221944)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 1, clone RP23-433P9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 221944)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 221944)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 19, 2003 this sequence version replaced gi:39930775.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L20488
                Center clone name: 433_P_9
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
```

```
            *        1    21153: contig of 21153 bp in length
            *    21154    21253: gap of 100 bp
            *    21254    26183: contig of 4930 bp in length
            *    26184    26283: gap of 100 bp
            *    26284    36809: contig of 10526 bp in length
            *    36810    36909: gap of 100 bp
            *    36910    42063: contig of 5154 bp in length
            *    42064    42163: gap of 100 bp
            *    42164    55640: contig of 13477 bp in length
            *    55641    55740: gap of 100 bp
            *    55741    71403: contig of 15663 bp in length
            *    71404    71503: gap of 100 bp
            *    71504    74916: contig of 3413 bp in length
            *    74917    75016: gap of 100 bp
            *    75017    78702: contig of 3686 bp in length
            *    78703    78802: gap of 100 bp
            *    78803    90026: contig of 11224 bp in length
            *    90027    90126: gap of 100 bp
            *    90127   112551: contig of 22425 bp in length
            *   112552   112651: gap of 100 bp
            *   112652   124702: contig of 12051 bp in length
            *   124703   124802: gap of 100 bp
            *   124803   128082: contig of 3280 bp in length
            *   128083   128182: gap of 100 bp
            *   128183   197115: contig of 68933 bp in length
            *   197116   197215: gap of 100 bp
            *   197216   221944: contig of 24729 bp in length.
FEATURES             Location/Qualifiers
     source          1. .221944
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /map="1"
                     /clone="RP23-433P9"
                     /clone_lib="RPCI-23 Female Mouse BAC"
ORIGIN

  Query Match             71.0%;  Score 22;  DB 2;  Length 221944;
  Best Local Similarity   83.3%;  Pred. No. 1.5e+02;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGAAC 30
              |||||| |||| || || || |||||||||
Db        426 TACAATGCAAAGATGGTGCACACAAGGAAC 455


RESULT 12
AC129026/c
LOCUS       AC129026              300680 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-43K10, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC129026
VERSION     AC129026.3  GI:25073698
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 300680)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 300680)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 300680)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23267514.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KANB
                Center clone name: CH230-43K10
            -------------- Summary Statistics
```

```
                Assembly program: Phrap; version 0.990329
                Consensus quality: 212742 bases at least Q40
                Consensus quality: 215738 bases at least Q30
                Consensus quality: 217892 bases at least Q20
                Estimated insert size: 213510; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
             --------------
             * NOTE: Estimated insert size may differ from sequence length
             *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 4 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *        1   258753: contig of 258753 bp in length
             *   258754   258853: gap of unknown length
             *   258854   287440: contig of 28587 bp in length
             *   287441   287540: gap of unknown length
             *   287541   298028: contig of 10488 bp in length
             *   298029   298128: gap of unknown length
             *   298129   300680: contig of 2552 bp in length.
FEATURES             Location/Qualifiers
     source          1. .300680
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-43K10"
     misc_feature    complement(256930. .257814)
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BH265728"
     misc_feature    287541. .288665
                     /note="wgs_end_extension
                     clone_end:T7"
ORIGIN

  Query Match             71.0%;  Score 22;  DB 2;  Length 300680;
  Best Local Similarity   83.3%;  Pred. No. 1.4e+02;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAACC 31
              |||||||||| |||||||||  || | |||
Db     229288 ACAATTCAAATATTGTTCAACAAAAGGACC 229259


RESULT 13
YSCMFAGB
LOCUS       YSCMFAGB                1165 bp    DNA     linear   PLN 27-APR-1993
DEFINITION  S.italicus mating factor alpha gene, complete cds.
ACCESSION   M55016 M35518 M35519
VERSION     M55016.1  GI:171944
KEYWORDS    mating factor alpha.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 1165)
  AUTHORS   Kitada,K. and Hishinuma,F.
  TITLE     Evidence for preferential multiplication of the internal unit in
            tandem repeats of the mating factor alpha genes in Saccharomyces
            yeasts
  JOURNAL   Curr. Genet. 13 (1), 1-5 (1988)
  MEDLINE   88194702
   PUBMED   2834101
COMMENT     Original source text: S.italicus, DNA.
FEATURES             Location/Qualifiers
     source          1. .1165
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
     CDS             293. .853
                     /codon_start=1
                     /product="mating factor alpha"
                     /protein_id="AAA34778.1"
                     /db_xref="GI:171945"
                     /translation="MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYLDL
                     EGDFDVAVLPFSNSTNNGLLFINTTIASIAAKEEGVSLDKREAEAWHWLQLKPGQPMY
                     KREADAEAWHWLQLKPGQPMYKREADAEAWHWLQLKPGQPMYKREADAEAWHWLQLKP
                     GQPMYKREADAEAWHWLQLKPGQPMY"
     mat_peptide     560. .598
                     /product="mating factor alpha"
     mat_peptide     623. .661
                     /product="mating factor alpha"
     mat_peptide     686. .724
                     /product="mating factor alpha"
     mat_peptide     749. .787
                     /product="mating factor alpha"
     mat_peptide     812. .850
                     /product="mating factor alpha"
ORIGIN

  Query Match             69.7%;  Score 21.6;  DB 8;  Length 1165;
  Best Local Similarity   85.7%;  Pred. No. 8.8e+02;
  Matches   24;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGA 28
              ||| ||||||  || ||||||||||||
Db        220 TACGATTCAAGAATAGTTCAAACAAGGA 247


RESULT 14
U41036/c
LOCUS       U41036                  37245 bp    DNA     linear   INV 21-NOV-2002
DEFINITION  Caenorhabditis elegans cosmid T14E8, complete sequence.
ACCESSION   U41036
VERSION     U41036.1  GI:1086751
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 37245)
  AUTHORS   Waterston,R.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
   PUBMED   9851916
REFERENCE   2  (bases 1 to 37245)
  AUTHORS   Hallsworth,K.
  TITLE     The sequence of C. elegans cosmid T14E8
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 37245)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-1995) Robert Waterston
REFERENCE   4  (bases 1 to 37245)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-2001) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   5  (bases 1 to 37245)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2001) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   6  (bases 1 to 37245)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2002) Department of Genetics, Washington
```

```
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   7  (bases 1 to 37245)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University
                  St. Louis , MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 IRQ, England
                  email: rw@nematode.wustl.edu and jes@sanger.ac.uk

            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate chemistry
            or covered by high quality data (i.e., phred quality >= 30); an
            attempt was made to resolve all sequencing problems, such as
            compressions and repeats; all regions were covered by sequence from
            more than one m13 subclone.

            For a graphical representation of this cosmid sequence and its
            analysis see:
            http://www.wormbase.org/db/seq/sequence?name=T14E8;class=Sequence

                        NEIGHBORING COSMID INFORMATION

            The 5' cosmid is F22A3, 200 bp overlap; the 3' cosmid is T28B4, 200
            bp overlap. Actual start of this cosmid is at base position 197 of
            T14E8; actual end is at 37245 of T14E8.

            NOTES:

            Coding seqences below are the result of integration and manual
            review of the following data : computer analysis using the program
            Genefinder (P. Green and L. Hillier, personal communication), the
            large scale EST projects of Yuji Kohara
            (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
            elegans ORFeome cloning project (http://worfdb.dfci.harvard.edu/),
            similarity to other proteins from BlastX analyses
            (http://blast.wustl.edu/), sequence conservation with C. briggsae
            using Jim Kent's WABA alignment program (Genome Research
            10:1115-1125, 2000), individual C. elegans GenBank submissions,
            and personal communications with C. elegans researchers. tRNAs
            are predicted using the program tRNAscan-SE (Lowe, T.M. and
            Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).
FEATURES              Location/Qualifiers
     source           1. .37245
                      /organism="Caenorhabditis elegans"
                      /mol_type="genomic DNA"
                      /strain="Bristol N2"
                      /db_xref="taxon:6239"
                      /chromosome="X"
                      /clone="T14E8"
     gene             5098. .7244
                      /gene="T14E8.2"
     CDS              join(5098. .5309,6255. .6324,6437. .6539,6612. .6784,
                      6855. .7244)
                      /gene="T14E8.2"
                      /standard_name="T14E8.2"
                      /note="contains similarity to Interpro domain IPR002172
                      (Low density lipoprotein-receptor, class A  domain)"
                      /codon_start=1
                      /product="Hypothetical protein T14E8.2"
                      /protein_id="AAA82384.1"
                      /db_xref="GI:1086752"
                      /db_xref="WormBase:T14E8.2"
                      /translation="MQSLEDKLSSPFVLCESQQTKITRGYDVLCLCCRRLKYVIVLIR
                      KLINLHFSASSRDRISNPVFFDVSESRMRRWFSVILLPFLVKIILAQELEEKNSLTRP
                      IHPRRHKRAEQVRPPLLSTSPFMLWKNDPKCMIPGRDMMACPKKDPNDPLEKLVCIEP
                      SALCDDHRDCHGAEDEDPHFCMPKKLEDAAMRRVQAEILVLVQNNRKLQPGVVVHIPD
                      KPDTGIVEPLRPGVFKVEGLGPLAKSKMYMNEKDDDDVNKSVEESDEEEENQESEEQE
                      DNDKAEKEEERRRQEEAARQHSKRRRTLRHLLQLKASYL"
     gene             16577. .22356
                      /gene="T14E8.1"
     CDS              join(16577. .16750,17388. .17469,17526. .17634,17682. .17885,
                      17931. .18181,18229. .18317,18367. .18582,18627. .18733,
                      18913. .19015,19063. .19182,19228. .19416,19465. .19586,
                      19633. .20025,20269. .20450,20503. .20755,20803. .20941,
                      21038. .21204,21760. .21912,22149. .22356)
                      /gene="T14E8.1"
                      /standard_name="T14E8.1a"
                      /note="contains similarity to Pfam domain PF00069
                      (Eukaryotic protein kinase domain); coded for by the
                      following C. elegans cDNAs: yk346e2.5, yk440e9.5,
                      yk192c3.3, yk272c1.3, yk346e2.3, yk206e9.3, yk30d4.3,
                      yk37f3.3, yk440e9.3, yk11b10.3, yk21h12.3, yk321e10.3,
                      yk7b2.5, yk11b10.5, yk14c10.3, yk21h12.5, yk30d4.5,
                      yk37f3.5, yk192c3.5, yk206e9.5, yk208e1.5, yk259f9.5,
                      yk272c1.5, yk321e10.5"
                      /codon_start=1
                      /product="Hypothetical protein T14E8.1a"
                      /protein_id="AAA82385.2"
                      /db_xref="GI:15011814"
                      /db_xref="WormBase:T14E8.1a"
                      /translation="MVLGSSQSSAKELTTQSSIFRFLVLLLCFTSATGGQINGKLLNG
                      NGVFVSRDELQNKNVLGVITGFDLLVVATHSRFQVFENNEERRTVGHVSLDMHPRTKF
                      LELKLFSKSEIFYCDESSCGLCTYSGVTSSCSTFMLNGDEPKIQEILSSSAVKIENLG
                      QIMLAISFKNEEDPDNPRTMILRYNAQDTGTVIPTAYHADSSFIHNNHALTGFEREGF
                      VYFVMTASQIFEPEVFLHDQSNNKVTVTKVIRFCATDQTADLASKISILVGCDQEFRN
                      ISSRGETAVYDHANDLINIVMFNHTSMNHLMCRFKMANIEKRFKTIWSTCQETSFSGS
                      TAKTTRCKYPQIFDQMKVKKGCLTYSRLDDESSPTLCVRYGRGDALDNCQLHTAKSNS
                      YRYGWLEDYNVLQGELMMRIPYPFFGIAESLITDGKSYFAAVSGEFDMSDVLRFSASE
                      SADIRPHWRTNISVVGKFSITKTKENQLLYTTVEGLQSLDISCKGLYPNCQTLRQGGW
                      EDPLECSWCADDNAQRTITSSEVSSCKNNLKHECPPSMRWIHKYNNNSGFTAVVDGFR
                      ALKNPKLNACGTNCVVTVVDSSSIQCDTNPDEVIGDSCKQVFLSGMIGDKNYSFPFDY
                      QQADRGTQTDVKNSQVDDKKGSSPGWKIAIAIISVMTIILIVAIIVYYMRNRFPRIKT
                      HVRPPIGQRIENEYDMGHMAGRQAQLAINGDNYVKVFRSMRPDLKVDFKNLRVDKLDP
                      IGQGHYGVVYKAMYSPSKSLEEKVVCKYLKEGKISEFYEEARTMSEFDHPNILKLIGV
                      ALDDSSHLPIIITEYMAKGDLKSFIENVENTIKMRDLFEFAFDIAKGMNYMHSKKFIH
                      RDLACRNCLLDEHLRVKIADFGLCRKVDIETELYVQMHERDLPVRWFPPEISEQGFGI
                      TSDIWSFGVVIWELFTRGSTPYSNMASWILILPWLKESETNRLRKPPYCPEKLYTDVM
                      LACWKANPAERPQFSDLVTIIPNVVKYMEGYDRSQLQAGYERVSSRFLSLSRHDPAFP
                      IYQNEMPNTPLLANCQNDTNDSKTLAELPSDSPSTSTAIPQSTPYQLLSECSETSV"
     CDS              join(16577. .16750,17388. .17469,17526. .17634,17682. .17885,
                      17931. .18181,18229. .18317,18367. .18582,18627. .18733,
                      18913. .19015,19063. .19182,19228. .19416,19465. .19586,
                      19633. .20025,20269. .20450,20503. .20755,20803. .20941,
                      21038. .21204,21760. .21901,22149. .22199)
                      /gene="T14E8.1"
                      /standard_name="T14E8.1b"
                      /note="contains similarity to Pfam domain PF00069
                      (Eukaryotic protein kinase domain); coded for by the
                      following C. elegans cDNAs: yk896b03.3, yk896b03.5"
                      /codon_start=1
                      /product="Hypothetical protein T14E8.1b"
                      /protein_id="AAL06055.1"
                      /db_xref="GI:15718663"
                      /db_xref="WormBase:T14E8.1b"
                      /translation="MVLGSSQSSAKELTTQSSIFRFLVLLLCFTSATGGQINGKLLNG
                      NGVFVSRDELQNKNVLGVITGFDLLVVATHSRFQVFENNEERRTVGHVSLDMHPRTKF
                      LELKLFSKSEIFYCDESSCGLCTYSGVTSSCSTFMLNGDEPKIQEILSSSAVKIENLG
                      QIMLAISFKNEEDPDNPRTMILRYNAQDTGTVIPTAYHADSSFIHNNHALTGFEREGF
                      VYFVMTASQIFEPEVFLHDQSNNKVTVTKVIRFCATDQTADLASKISILVGCDQEFRN
                      ISSRGETAVYDHANDLINIVMFNHTSMNHLMCRFKMANIEKRFKTIWSTCQETSFSGS
                      TAKTTRCKYPQIFDQMKVKKGCLTYSRLDDESSPTLCVRYGRGDALDNCQLHTAKSNS
                      YRYGWLEDYNVLQGELMMRIPYPFFGIAESLITDGKSYFAAVSGEFDMSDVLRFSASE
                      SADIRPHWRTNISVVGKFSITKTKENQLLYTTVEGLQSLDISCKGLYPNCQTLRQGGW
                      EDPLECSWCADDNAQRTITSSEVSSCKNNLKHECPPSMRWIHKYNNNSGFTAVVDGFR
```

```
                     ALKNPKLNACGTNCVVTVVDSSSIQCDTNPDEVIGDSCKQVFLSGMIGDKNYSFPFDY
                     QQADRGTQTDVKNSQVDDKKGSSPGWKIAIAIISVMTIILIVAIIVYYMRNRFPRIKT
                     HVRPPIGQRIENEYDMGHMAGRQAQLAINGDNYVKVFRSMRPDLKVDFKNLRVDKLDP
                     IGQGHYGVVYKAMYSPSKSLEEKVVCKYLKEGKISEFYEEARTMSEFDHPNILKLIGV
                     ALDDSSHLPIIITEYMAKGDLKSFIENVENTIKMRDLFEFAFDIAKGMNYMHSKKFIH
                     RDLACRNCLLDEHLRVKIADFGLCRKVDIETELYVQMHERDLPVRWFPPEISEQGFGI
                     TSDIWSFGVVIWELFTRGSTPYSNMASWILILPWLKESETNRLRKPPYCPEKLYTDVM
                     LACWKANPAERPQFSDLVTIIPNVVKYMEGYDRSQLQAGYERVSELVTSRPSISNLSE
                     "
     gene            complement(23380. .32779)
                     /gene="T14E8.3"
     CDS             complement(join(23380. .23408,23458. .23704,23753. .23932,
                     23988. .24467,24513. .24639,25078. .25320,25371. .25444,
                     25486. .25604,25662. .25726,27723. .27849,27904. .28022,

  Query Match             69.7%;  Score 21.6;  DB 3;  Length 37245;
  Best Local Similarity   85.7%;  Pred. No. 3.5e+02;
  Matches   24;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAA 29
              | ||||||||||||| || ||| ||||||
Db      16418 AAAATTCAAACATTTTTAAAAGAAGGAA 16391


RESULT 15
AC024750/c
LOCUS       AC024750               41335 bp    DNA     linear   INV 15-JUL-2003
DEFINITION  Caenorhabditis elegans cosmid Y17G9A, complete sequence.
ACCESSION   AC024750
VERSION     AC024750.1  GI:7140301
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 41335)
  AUTHORS   Wilson,R.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
   PUBMED   9851916
REFERENCE   2  (bases 1 to 41335)
  AUTHORS   Fulton,B., Courtney,L. and Ames,M.
  TITLE     The sequence of C. elegans cosmid Y17G9A
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 41335)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 41335)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 41335)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAY-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   6  (bases 1 to 41335)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   7  (bases 1 to 41335)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-2003) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University
                  St. Louis , MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 IRQ, England
                  email: submissions@watson.wustl.edu and jes@sanger.ac.uk

            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate chemistry
            or covered by high quality data (i.e., phred quality >= 30); an
            attempt was made to resolve all sequencing problems, such as
            compressions and repeats; all regions were covered by sequence from
            more than one m13 subclone.

            For a graphical representation of this clone sequence and its
            analysis see:
            http://www.wormbase.org/db/seq/sequence?name=Y17G9A;class=Sequence

                        NEIGHBORING CLONE INFORMATION

            The 5' clone is H06H21, 7000 bp overlap; the 3' clone is T08B6, 200
            bp overlap. Actual start of this clone is at base position 18001 of
            K08F11.

            NOTES:

            Coding seqences below are the result of integration and manual
            review of the following data : computer analysis using the program
            Genefinder (P. Green and L. Hillier, personal communication), the
            large scale EST projects of Yuji Kohara
            (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
            elegans ORFeome cloning project (http://worfdb.dfci.harvard.edu/),
            similarity to other proteins from BlastX analyses
            (http://blast.wustl.edu/), sequence conservation with C. briggsae
            using Jim Kent's WABA alignment program (Genome Research
            10:1115-1125, 2000), individual C. elegans GenBank submissions,
            and personal communications with C. elegans researchers. tRNAs
            are predicted using the program tRNAscan-SE (Lowe, T.M. and
            Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).
FEATURES             Location/Qualifiers
     source          1. .41335
                     /organism="Caenorhabditis elegans"
                     /mol_type="genomic DNA"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="IV"
                     /clone="Y17G9A"
     gene            complement(19722. .21548)
                     /gene="Y17G9A.2"
     CDS             complement(join(19722. .19941,19985. .20194,20494. .20638,
                     20684. .20796,21514. .21548))
                     /gene="Y17G9A.2"
                     /standard_name="Y17G9A.2"
                     /note="coded for by the following C. elegans cDNAs:
                     yk726c11.3, yk726c11.5"
                     /codon_start=1
                     /product="Hypothetical protein Y17G9A.2"
                     /protein_id="AAF60437.2"
                     /db_xref="GI:21166341"
                     /db_xref="WormBase:Y17G9A.2"
                     /translation="MQQNRRLDDSSCPSQTLVVLTLLSLFLCAFTAFGSAFLLYQLMD
                     STSCKDPRLPLKMDYVWPGNVQKIIDYVIENDMHVTNNDRTFIIRTELMRFKTTDEFY
                     IIVYDHTPESNNHAFYGNKDQYIVSINPGKSNVVVYRSRFWKSSQDSAYNKVMREVDA
                     LKENGIQFETDYSHLPKQLSRNFTDSNFIGIISNDLNPKVRGANNWARYWGPGYFEEI
                     PVFDGTTKQEKGSKITVVVGYK"
```

```
gene            complement(22168. .23950)
                /gene="Y17G9A.3"
CDS             complement(join(22168. .22378,22896. .23105,23438. .23582,
                23625. .23728,23916. .23950))
                /gene="Y17G9A.3"
                /standard_name="Y17G9A.3"
                /note="contains similarity to Pfam domain PF05075 (Protein
                of unknown function (DUF684))"
                /codon_start=1
                /product="Hypothetical protein Y17G9A.3"
                /protein_id="AAF60439.1"
                /db_xref="GI:7331751"
                /db_xref="WormBase:Y17G9A.3"
                /translation="MQKNRPLKDFTSSNQGLVLILVGLWLFALSSFGLTFFSLQRIAS
                VESSKLPLKLKYIWPKNVKKIVDYVIENDMHVTNDDRAVLIKTELMRFDTTDVFYIMV
                YDNKGETESNSFYGNAEQYITSYNPGKCNIVIYRSHYWRSTSKRAKNNIENEVKIICA
                EELKKQPDGLENLIKNIEDARFIGLIEKNRAVSIRSANSFGRARGPGWWNKCLEVYPD
                TKKLTGKQFILIVGFK"
gene            complement(24532. .26683)
                /gene="Y17G9A.4"
CDS             complement(join(24532. .24754,25107. .25239,25290. .25393,
                25700. .25891,25939. .26148,26287. .26431,26479. .26591,
                26649. .26683))
                /gene="Y17G9A.4"
                /standard_name="Y17G9A.4"
                /note="contains similarity to Interpro domain IPR002040
                (Tachykinin)"
                /codon_start=1
                /product="Hypothetical protein Y17G9A.4"
                /protein_id="AAF60440.1"
                /db_xref="GI:7331752"
                /db_xref="WormBase:Y17G9A.4"
                /translation="MSQNRRTTDCENSPKQCLLLILISLFLCGLIAFGLTYFIQKWIE
                TENVETSGIPLKMDYIWPQNVKKIVDYVIENDMHVTNDDRAVLIKTELMRFNTSDVFY
                IMVYDNKGETESNSFYGNAEQYITSYNPGKSNIVIYRSHYWRSTTENAKDNIENEVKI
                ICAEGLNKQPDYNDVPVKLSENLENSRFVGLMGKDLDVSIRSANSFGRVWGPGWWDSI
                DEVYPDSKKQTGKHSSQHFLLILAFLWLCALISFGVTLIIIQWIEKVKSPRLPLEMDY
                IWPKNVKKIVDYVIENDMHVTNYDRAVLIKTELMRQPDISISLEKLASEIENSRFIGL
                FEKYQEVAIESANSFNGRGWGPGWWNKCLEVYPDSKKKTGKQFTLIAGFQ"
tRNA            27032. .27098
                /product="tRNA-His"
                /standard_name="Y17G9A.t1"
                /note="C. elegans predicted tRNA"
                /db_xref="WormBase:Y17G9A.t1"
gene            27462. .29380
                /gene="str-173"
CDS             join(27462. .27659,27933. .28020,28067. .28145,28355. .28489,
                28759. .29032,29078. .29255,29313. .29380)
                /gene="str-173"
                /standard_name="Y17G9A.1"
                /note="contains similarity to Pfam domain PF01461 (7TM
                chemoreceptor); coded for by the following C. elegans
                cDNAs: OSTF078B3_1"
                /codon_start=1
                /product="Seven tm receptor protein 173"
                /protein_id="AAF60438.2"
                /db_xref="GI:32699204"
                /db_xref="WormBase:Y17G9A.1"
                /translation="MTPVTFANIKTVIQFISVFISIFINALSIYLIVTKSPKIMGTYR
                HLMIYFCGCSMVFSVLDVIVQPSIETYKSAFLMVVDVKNRSLNSWSAVLFMNLLCGCI
                GVTVYGIAIHFIYRLLALERKGRLRYFNCQYLPIWFSIPILAGTIWYIVTENIFGMNP
                VNTEYIRNPVKDSFNMKMEDCVYGAAVFYPIDKNGIKTISWISFFGLGCYLTLLTIPF
                ISILVAGAKSWNHVQKLLEHGESEFARNLQMQLYKALIAQTLIPVVFFFTPFGCVFVL
                PAFEIECQFMSAPITLVFAIYPAIDPLPNLFFVDYYRNPISNFLRKVWCRKPKVEAAT
                INSTL"
gene            complement(30257. .32936)
                /gene="str-172"
CDS             complement(join(30257. .30342,30391. .30568,30620. .30893,
                31018. .31276,32600. .32687,32739. .32936))
                /gene="str-172"
                /standard_name="Y17G9A.5"
                /note="contains similarity to Pfam domain PF01461 (7TM
                chemoreceptor)"
                /codon_start=1
                /product="Seven tm receptor protein 172"
                /protein_id="AAF60441.1"
                /db_xref="GI:7331753"
                /db_xref="WormBase:Y17G9A.5"
                /translation="MHATTFESIKFVFQVTSVIFSWLINSFLIYLILTKSTSKMGNYR
                HLMIYFCCFSILFSTLDIIVQPNIHTYKSAFFMIMDYKNRGIPLWMAEVLICLMCGCF
```

```
  Query Match             69.7%;  Score 21.6;  DB 3;  Length 41335;
  Best Local Similarity   85.7%;  Pred. No. 3.4e+02;
  Matches   24;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGA 28
              || ||||| | ||||||| |||||||||
Db      30704 TATAATTCGAGCATTGTTGAAACAAGGA 30677


Search completed: July 20, 2004, 00:54:45
Job time : 775.426 secs
```

This page blank (usp.)

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 00:16:33 ; Search time 35.4098 Seconds
(without alignments)
470.167 Million cell updates/sec

Title: US-09-831-272-2
Perfect score: 30
Sequence: 1 cacacttaatttgaccgagtaacattcgcc 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	19.4	64.7	61	1	US-08-207-547A-30	Sequence 30, Appl
	2	19.4	64.7	61	1	US-08-215-082-30	Sequence 30, Appl
	3	19.4	64.7	61	2	US-08-702-652-30	Sequence 30, Appl
	4	17.6	58.7	521	4	US-09-976-594-1112	Sequence 1112, Ap
c	5	17.2	57.3	768	5	PCT-US95-10973A-29	Sequence 29, Appl
c	6	17.2	57.3	768	5	PCT-US95-10973A-30	Sequence 30, Appl
c	7	17.2	57.3	771	5	PCT-US95-10973A-59	Sequence 59, Appl
c	8	17.2	57.3	804	2	US-08-356-161-3	Sequence 3, Appli
c	9	17.2	57.3	804	2	US-08-356-161-4	Sequence 4, Appli
c	10	17.2	57.3	804	2	US-08-356-161-5	Sequence 5, Appli
c	11	17.2	57.3	804	2	US-08-356-161-6	Sequence 6, Appli
c	12	17.2	57.3	804	2	US-08-356-161-7	Sequence 7, Appli
c	13	17.2	57.3	804	3	US-08-718-904-19	Sequence 19, Appl
c	14	17.2	57.3	804	3	US-08-718-904-20	Sequence 20, Appl
c	15	17.2	57.3	804	3	US-08-718-904-21	Sequence 21, Appl
c	16	17.2	57.3	804	3	US-08-718-904-22	Sequence 22, Appl
c	17	17.2	57.3	804	3	US-08-718-904-23	Sequence 23, Appl
c	18	17.2	57.3	804	4	US-09-449-249-19	Sequence 19, Appl
c	19	17.2	57.3	804	4	US-09-449-249-20	Sequence 20, Appl
c	20	17.2	57.3	804	4	US-09-449-249-21	Sequence 21, Appl
c	21	17.2	57.3	804	4	US-09-449-249-22	Sequence 22, Appl
c	22	17.2	57.3	804	4	US-09-449-249-23	Sequence 23, Appl
c	23	17.2	57.3	804	5	PCT-US93-05702-3	Sequence 3, Appli
c	24	17.2	57.3	804	5	PCT-US93-05702-4	Sequence 4, Appli
c	25	17.2	57.3	804	5	PCT-US93-05702-5	Sequence 5, Appli
c	26	17.2	57.3	804	5	PCT-US93-05702-6	Sequence 6, Appli
c	27	17.2	57.3	804	5	PCT-US93-05702-7	Sequence 7, Appli
c	28	17.2	57.3	804	5	PCT-US95-10973A-3	Sequence 3, Appli
c	29	17.2	57.3	804	5	PCT-US95-10973A-4	Sequence 4, Appli
c	30	17.2	57.3	804	5	PCT-US95-10973A-5	Sequence 5, Appli
c	31	17.2	57.3	804	5	PCT-US95-10973A-6	Sequence 6, Appli
c	32	17.2	57.3	804	5	PCT-US95-10973A-7	Sequence 7, Appli
c	33	17.2	57.3	1167	5	PCT-US95-10973A-57	Sequence 57, Appl
c	34	17.2	57.3	1212	5	PCT-US95-10973A-31	Sequence 31, Appl
c	35	17.2	57.3	1230	2	US-08-356-161-12	Sequence 12, Appl
c	36	17.2	57.3	1230	2	US-08-356-161-13	Sequence 13, Appl
c	37	17.2	57.3	1230	3	US-08-718-904-53	Sequence 53, Appl
c	38	17.2	57.3	1230	4	US-09-449-249-53	Sequence 53, Appl
c	39	17.2	57.3	1230	5	PCT-US93-05702-12	Sequence 12, Appl
c	40	17.2	57.3	1230	5	PCT-US93-05702-13	Sequence 13, Appl
c	41	17.2	57.3	1230	5	PCT-US95-10973A-12	Sequence 12, Appl
c	42	17.2	57.3	1230	5	PCT-US95-10973A-13	Sequence 13, Appl
c	43	17.2	57.3	1251	3	US-08-718-904-74	Sequence 74, Appl
c	44	17.2	57.3	1251	4	US-09-449-249-74	Sequence 74, Appl
c	45	17.2	57.3	1260	3	US-08-718-904-72	Sequence 72, Appl

ALIGNMENTS

RESULT 1
US-08-207-547A-30
; Sequence 30, Application US/08207547A
; Patent No. 5624824
; GENERAL INFORMATION:
; APPLICANT: Yuan, Yan
; APPLICANT: Guerrier-Takada, Cecilia
; APPLICANT: Altman, Sidney
; APPLICANT: Liu, Fenyong
; TITLE OF INVENTION: Targeted Cleavage of RNA Using
; TITLE OF INVENTION: Eukaryotic Ribonuclease P and External Guide Sequence
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/207,547A
; FILING DATE: 07-MAR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/03961
; FILING DATE: 28-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/054,892
; FILING DATE: 29-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/931,837
; FILING DATE: 18-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/875,099
; FILING DATE: 28-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/568,834
; FILING DATE: 17-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/328,368
; FILING DATE: 24-MAR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284

```
;      REFERENCE/DOCKET NUMBER:  YU100CIP(4)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404)-873-8794
;      TELEFAX:  (404)-873-8795
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  61 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  tRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-207-547A-30

  Query Match             64.7%;  Score 19.4;  DB 1;  Length 61;
  Best Local Similarity   58.6%;  Pred. No. 2.4;
  Matches   17;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          2 ACACTTAATTTGACCGAGTAACATTCGCC 30
              ||||: ||:::|  |||| ||  ::||||
Db         16 ACACUAAAUUUGCACGAGGAAGGUUCGCC 44


RESULT 2
US-08-215-082-30
; Sequence 30, Application US/08215082
; Patent No. 5728521
;  GENERAL INFORMATION:
;    APPLICANT:  Yuan, Yan
;    APPLICANT:  Guerrier-Takada, Cecilia
;    APPLICANT:  Altman, Sidney
;    APPLICANT:  Liu, Fenyong
;    TITLE OF INVENTION:  Targeted Cleavage of RNA Using
;    TITLE OF INVENTION:  Eukaryotic Ribonuclease P and External Guide Sequence
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  1100 Peachtree Street, Suite 2800
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-4530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/215,082
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  YU100CIP(4)
;      FILING DATE:  07-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/03961
;      FILING DATE:  28-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/054,892
;      FILING DATE:  29-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/931,837
;      FILING DATE:  18-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/875,099
;      FILING DATE:  28-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/568,834
;      FILING DATE:  17-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/328,368
;      FILING DATE:  24-MAR-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  YU100CIP(5)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404)-815-6508
;      TELEFAX:  (404)-815-6555
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  61 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  tRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-215-082-30

  Query Match             64.7%;  Score 19.4;  DB 1;  Length 61;
  Best Local Similarity   58.6%;  Pred. No. 2.4;
  Matches   17;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          2 ACACTTAATTTGACCGAGTAACATTCGCC 30
              ||||: ||:::|  |||| ||  ::||||
Db         16 ACACUAAAUUUGCACGAGGAAGGUUCGCC 44


RESULT 3
US-08-702-652-30
; Sequence 30, Application US/08702652
; Patent No. 5869248
;  GENERAL INFORMATION:
;    APPLICANT:  Yan Yuan, Cecilia Guerrier-Takada, and
;    APPLICANT:  Sidney Altman
;    TITLE OF INVENTION:  TARGETED CLEAVAGE OF RNA USING
;    TITLE OF INVENTION:  RIBONUCLEASE P TARGETING AND CLEAVAGE SEQUENCES
;    NUMBER OF SEQUENCES:  65
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center
;      STREET:  1201 West Peachtree Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/702,652
;      FILING DATE:  No. 5869248ember 6, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/02816
;      FILING DATE:  March 7, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/207,547
;      FILING DATE:  March 7, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  YU112
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404)-873-8794
;      TELEFAX:  (404)-873-8795
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  61 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  tRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-702-652-30

  Query Match             64.7%;  Score 19.4;  DB 2;  Length 61;
  Best Local Similarity   58.6%;  Pred. No. 2.4;
  Matches   17;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          2 ACACTTAATTTGACCGAGTAACATTCGCC 30
              |||| ||:::|  |||| ||  ::||||
Db         16 ACACUAAAUUUGCACGAGGAAGGUUCGCC 44


RESULT 4
US-09-976-594-1112
; Sequence 1112, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
;  APPLICANT: Furness, Michael
;  APPLICANT:  Buchbinder, Jenny
;  TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
;  FILE REFERENCE: PA-0041 US
;  CURRENT APPLICATION NUMBER: US/09/976,594
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: 60/240,409
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 1143
;  SOFTWARE: PERL Program
; SEQ ID NO 1112
;    LENGTH: 521
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    OTHER INFORMATION: Incyte ID No. 6673549 1040914.1
US-09-976-594-1112

  Query Match             58.7%;  Score 17.6;  DB 4;  Length 521;
  Best Local Similarity   83.3%;  Pred. No. 26;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 ACACTTAATTTGACCGAGTAACAT 25
              ||| ||| || ||| |||||||||
Db        245 ACAATTATTTGGACTGAGTAACAT 268


RESULT 5
PCT-US95-10973A-29/c
; Sequence 29, Application PC/TUS9510973A
;  GENERAL INFORMATION:
;     APPLICANT:  Prizm Pharmaceuticals, Inc.
;     TITLE OF INVENTION:  CONJUGATES OF VASCULAR ENDOTHELIAL GROWTH FACTOR WITH TARGET
;     NUMBER OF SEQUENCES:  107
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SEED and BERRY
;       STREET:  6300 Columbia Center, 701 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98104-7092
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/10973A
;       FILING DATE:  29-AUG-1995
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Nottenburg, Carol
;       REGISTRATION NUMBER:  39,317
;       REFERENCE/DOCKET NUMBER:  760100.413PC
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  29:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  768 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  both
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  4..768
;       OTHER INFORMATION:  /product= "SAP CYS +4"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  7..768
;       OTHER INFORMATION:  /product= "mature SAP CYS +4"
PCT-US95-10973A-29

  Query Match             57.3%;  Score 17.2;  DB 5;  Length 768;
  Best Local Similarity   86.4%;  Pred. No. 44;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          7 TAATTTGACCGAGTAACATTCG 28
              |||| || |||| |||||||||
Db        272 TAATATGCCCGATTAACATTCG 251


RESULT 6
PCT-US95-10973A-30/c
; Sequence 30, Application PC/TUS9510973A
;  GENERAL INFORMATION:
;     APPLICANT:  Prizm Pharmaceuticals, Inc.
;     TITLE OF INVENTION:  CONJUGATES OF VASCULAR ENDOTHELIAL GROWTH FACTOR WITH TARGET
;     NUMBER OF SEQUENCES:  107
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SEED and BERRY
;       STREET:  6300 Columbia Center, 701 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98104-7092
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/10973A
;       FILING DATE:  29-AUG-1995
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Nottenburg, Carol
;       REGISTRATION NUMBER:  39,317
;       REFERENCE/DOCKET NUMBER:  760100.413PC
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  30:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  768 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  both
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  4..768
;       OTHER INFORMATION:  /product= "SAP CYS +10"
;     FEATURE:
```

```
;       NAME/KEY:  CDS
;       LOCATION:  7..768
;       OTHER INFORMATION:  /product= "mature SAP CYS +10"
PCT-US95-10973A-30

  Query Match             57.3%;  Score 17.2;  DB 5;  Length 768;
  Best Local Similarity   86.4%;  Pred. No. 44;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          7 TAATTTGACCGAGTAACATTCG 28
              |||| || |||| |||||||||
Db        272 TAATATGCCCGATTAACATTCG 251


RESULT 7
PCT-US95-10973A-59/c
; Sequence 59, Application PC/TUS9510973A
;  GENERAL INFORMATION:
;    APPLICANT:  Prizm Pharmaceuticals, Inc.
;    TITLE OF INVENTION:  CONJUGATES OF VASCULAR ENDOTHELIAL GROWTH FACTOR WITH TARGET
;    NUMBER OF SEQUENCES:  107
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10973A
;      FILING DATE:  29-AUG-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nottenburg, Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  760100.413PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  771 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  4..771
;      OTHER INFORMATION:  /product= "SAP CYS -1"
PCT-US95-10973A-59

  Query Match             57.3%;  Score 17.2;  DB 5;  Length 771;
  Best Local Similarity   86.4%;  Pred. No. 44;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          7 TAATTTGACCGAGTAACATTCG 28
              |||| || |||| |||||||||
Db        275 TAATATGCCCGATTAACATTCG 254


RESULT 8
US-08-356-161-3/c
; Sequence 3, Application US/08356161
; Patent No. 5916772
;  GENERAL INFORMATION:
;    APPLICANT:  Lappi, Douglas A.
;    APPLICANT:  Barthelemy, Isabel
;    APPLICANT:  Baird, J. Andrew
;    APPLICANT:  Sosnowski, Barbara A.
;    TITLE OF INVENTION:  RECOMBINANT PRODUCTION OF
;    TITLE OF INVENTION:  SAPORIN-CONTAINING PROTEINS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  701 Fifth Avenue, 6300 Columbia Center
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/356,161
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/356,161
;      FILING DATE:  13-APR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5916772tenburg, Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  760100.404US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  804 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..804
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..804
;      OTHER INFORMATION:  /note= "Nucleotide sequence
;      OTHER INFORMATION:  corresponding to the clone M13 mp18-G4 in Example 1.B.2."
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  46..804
;      OTHER INFORMATION:  /product= ""Saporin""
US-08-356-161-3

  Query Match             57.3%;  Score 17.2;  DB 2;  Length 804;
  Best Local Similarity   86.4%;  Pred. No. 44;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          7 TAATTTGACCGAGTAACATTCG 28
              |||| || |||| |||||||||
Db        311 TAATATGCCCGATTAACATTCG 290


RESULT 9
US-08-356-161-4/c
; Sequence 4, Application US/08356161
; Patent No. 5916772
;  GENERAL INFORMATION:
;    APPLICANT:  Lappi, Douglas A.
;    APPLICANT:  Barthelemy, Isabel
;    APPLICANT:  Baird, J. Andrew
;    APPLICANT:  Sosnowski, Barbara A.
;    TITLE OF INVENTION:  RECOMBINANT PRODUCTION OF
;    TITLE OF INVENTION:  SAPORIN-CONTAINING PROTEINS
;    NUMBER OF SEQUENCES:  15
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  701 Fifth Avenue, 6300 Columbia Center
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/356,161
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/356,161
;      FILING DATE:  13-APR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5916772tenburg, Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  760100.404US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  804 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..804
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..804
;      OTHER INFORMATION:  /note= "Nucleotide sequence
;      OTHER INFORMATION:  corresponding to the clone M13 mp18-G1 in Example 1.B.2."
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  46..804
;      OTHER INFORMATION:  /product= "Saporin"
US-08-356-161-4

  Query Match             57.3%;  Score 17.2;  DB 2;  Length 804;
  Best Local Similarity   86.4%;  Pred. No. 44;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          7 TAATTTGACCGAGTAACATTCG 28
              |||| || |||| |||||||||
Db        311 TAATATGCCCGATTAACATTCG 290


RESULT 10
US-08-356-161-5/c
; Sequence 5, Application US/08356161
; Patent No. 5916772
;  GENERAL INFORMATION:
;    APPLICANT:  Lappi, Douglas A.
;    APPLICANT:  Barthelemy, Isabel
;    APPLICANT:  Baird, J. Andrew
;    APPLICANT:  Sosnowski, Barbara A.
;    TITLE OF INVENTION:  RECOMBINANT PRODUCTION OF
;    TITLE OF INVENTION:  SAPORIN-CONTAINING PROTEINS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  701 Fifth Avenue, 6300 Columbia Center
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/356,161
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/356,161
;      FILING DATE:  13-APR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5916772tenburg, Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  760100.404US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  804 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..804
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..804
;      OTHER INFORMATION:  /note= "Nucleotide sequence
;      OTHER INFORMATION:  corresponding to the clone M13 mp18-G2 in Example 1.B.2."
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  46..804
;      OTHER INFORMATION:  /product= "Saporin"
US-08-356-161-5

  Query Match             57.3%;  Score 17.2;  DB 2;  Length 804;
  Best Local Similarity   86.4%;  Pred. No. 44;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          7 TAATTTGACCGAGTAACATTCG 28
              |||| || |||| |||||||||
Db        311 TAATATGCCCGATTAACATTCG 290


RESULT 11
US-08-356-161-6/c
; Sequence 6, Application US/08356161
; Patent No. 5916772
;  GENERAL INFORMATION:
;    APPLICANT:  Lappi, Douglas A.
;    APPLICANT:  Barthelemy, Isabel
;    APPLICANT:  Baird, J. Andrew
;    APPLICANT:  Sosnowski, Barbara A.
;    TITLE OF INVENTION:  RECOMBINANT PRODUCTION OF
;    TITLE OF INVENTION:  SAPORIN-CONTAINING PROTEINS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  701 Fifth Avenue, 6300 Columbia Center
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/356,161
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/356,161
;       FILING DATE:  13-APR-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  No. 5916772tenburg, Carol
;       REGISTRATION NUMBER:  39,317
;       REFERENCE/DOCKET NUMBER:  760100.404US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  804 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..804
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1..804
;       OTHER INFORMATION:  /note= "Nucleotide sequence
;       OTHER INFORMATION:  corresponding to the clone M13 mp18-G7 in Example 1.B.2."
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  46..804
;       OTHER INFORMATION:  /product= "Saporin"
US-08-356-161-6

  Query Match             57.3%;  Score 17.2;  DB 2;  Length 804;
  Best Local Similarity   86.4%;  Pred. No. 44;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          7 TAATTTGACCGAGTAACATTCG 28
              |||| || |||| |||||||||
Db        311 TAATATGCCCGATTAACATTCG 290


RESULT 12
US-08-356-161-7/c
; Sequence 7, Application US/08356161
; Patent No. 5916772
;  GENERAL INFORMATION:
;    APPLICANT:  Lappi, Douglas A.
;    APPLICANT:  Barthelemy, Isabel
;    APPLICANT:  Baird, J. Andrew
;    APPLICANT:  Sosnowski, Barbara A.
;    TITLE OF INVENTION:  RECOMBINANT PRODUCTION OF
;    TITLE OF INVENTION:  SAPORIN-CONTAINING PROTEINS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  701 Fifth Avenue, 6300 Columbia Center
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/356,161
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/356,161
;      FILING DATE:  13-APR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5916772tenburg, Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  760100.404US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  804 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..804
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..804
;      OTHER INFORMATION:  /note= "Nucleotide sequence
;      OTHER INFORMATION:  corresponding to the clone M13 mp18-G9 in Example 1.B.2."
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  46..804
;      OTHER INFORMATION:  /product= "Saporin"
US-08-356-161-7

  Query Match             57.3%;  Score 17.2;  DB 2;  Length 804;
  Best Local Similarity   86.4%;  Pred. No. 44;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          7 TAATTTGACCGAGTAACATTCG 28
              |||| || |||| |||||||||
Db        311 TAATATGCCCGATTAACATTCG 290


RESULT 13
US-08-718-904-19/c
; Sequence 19, Application US/08718904
; Patent No. 6037329
;  GENERAL INFORMATION:
;    APPLICANT:  Baird, J. Andrew
;    APPLICANT:  Chandler, Lois Ann
;    APPLICANT:  Sosnowski, Barbara A.
;    TITLE OF INVENTION:  COMPOSITIONS CONTAINING NUCLEIC ACIDS AND LIGANDS FOR THERAPE
;    NUMBER OF SEQUENCES:  128
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/718,904
;      FILING DATE:  24-SEP-1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6037329tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  760100.415C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
```

```
;       TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  804 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..804
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..804
;      OTHER INFORMATION:  /note= "Nucleotide sequence
;      OTHER INFORMATION:  corresponding to the clone M13 mp18-G4 in Example I.B.2."
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  46..804
;      OTHER INFORMATION:  /product= ""Saporin""
US-08-718-904-19

  Query Match             57.3%;  Score 17.2;  DB 3;  Length 804;
  Best Local Similarity   86.4%;  Pred. No. 44;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          7 TAATTTGACCGAGTAACATTCG 28
              |||| || |||| |||||||||
Db        311 TAATATGCCCGATTAACATTCG 290


RESULT 14
US-08-718-904-20/c
; Sequence 20, Application US/08718904
; Patent No. 6037329
;  GENERAL INFORMATION:
;    APPLICANT:  Baird, J. Andrew
;    APPLICANT:  Chandler, Lois Ann
;    APPLICANT:  Sosnowski, Barbara A.
;    TITLE OF INVENTION:  COMPOSITIONS CONTAINING NUCLEIC ACIDS AND LIGANDS FOR THERAPE
;    NUMBER OF SEQUENCES:  128
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/718,904
;      FILING DATE:  24-SEP-1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6037329tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  760100.415C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  804 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..804
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..804
;      OTHER INFORMATION:  /note= "Nucleotide sequence
;      OTHER INFORMATION:  corresponding to the clone M13 mp18-G1 in Example I.B.2."
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  46..804
;      OTHER INFORMATION:  /product= "Saporin"
US-08-718-904-20

  Query Match             57.3%;  Score 17.2;  DB 3;  Length 804;
  Best Local Similarity   86.4%;  Pred. No. 44;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          7 TAATTTGACCGAGTAACATTCG 28
              |||| || |||| |||||||||
Db        311 TAATATGCCCGATTAACATTCG 290


RESULT 15
US-08-718-904-21/c
; Sequence 21, Application US/08718904
; Patent No. 6037329
;  GENERAL INFORMATION:
;    APPLICANT:  Baird, J. Andrew
;    APPLICANT:  Chandler, Lois Ann
;    APPLICANT:  Sosnowski, Barbara A.
;    TITLE OF INVENTION:  COMPOSITIONS CONTAINING NUCLEIC ACIDS AND LIGANDS FOR THERAPE
;    NUMBER OF SEQUENCES:  128
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/718,904
;      FILING DATE:  24-SEP-1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6037329tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  760100.415C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  804 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..804
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..804
;      OTHER INFORMATION:  /note= "Nucleotide sequence
;      OTHER INFORMATION:  corresponding to the clone M13 mp18-G2 in Example I.B.2."
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  46..804
;      OTHER INFORMATION:  /product= "Saporin"
```

US-08-718-904-21

Query Match 57.3%; Score 17.2; DB 3; Length 804;
Best Local Similarity 86.4%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          7 TAATTTGACCGAGTAACATTCG 28
              |||| || |||| |||||||||
Db        311 TAATATGCCCGATTAACATTCG 290
```

Search completed: July 20, 2004, 00:55:57
Job time : 36.4098 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 00:23:29 ; Search time 166.721 Seconds
(without alignments)
877.864 Million cell updates/sec

Title: US-09-831-272-2
Perfect score: 30
Sequence: 1 cacacttaatttgaccgagtaacattcgcc 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3190992 seqs, 2439311697 residues

Total number of hits satisfying chosen parameters: 6381984

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	19.4	64.7	3673778	15	US-10-312-841-2	Sequence 2, Appli
c	2	19	63.3	684707	16	US-10-398-221-9	Sequence 9, Appli
c	3	19	63.3	3011208	16	US-10-398-221-2058	Sequence 2058, Ap
	4	18.6	62.0	180	13	US-10-085-783A-42095	Sequence 42095, A
	5	18.6	62.0	180	16	US-10-242-535A-42095	Sequence 42095, A
c	6	18.6	62.0	612	13	US-10-027-632-217354	Sequence 217354,
c	7	18.6	62.0	612	13	US-10-027-632-217355	Sequence 217355,
c	8	18.6	62.0	612	16	US-10-027-632-217354	Sequence 217354,
c	9	18.6	62.0	612	16	US-10-027-632-217355	Sequence 217355,
	10	18.6	62.0	868	16	US-10-350-923B-5	Sequence 5, Appli
	11	18.6	62.0	1033	15	US-10-198-846-9689	Sequence 9689, Ap
c	12	18.4	61.3	7500	15	US-10-311-455-1088	Sequence 1088, Ap
	13	18.4	61.3	33112	16	US-10-429-873A-3	Sequence 3, Appli
	14	18.4	61.3	39001	16	US-10-199-199-14	Sequence 14, Appl
	15	18.4	61.3	65608	9	US-09-962-436-292	Sequence 292, App
	16	18.4	61.3	65608	9	US-09-962-832-119	Sequence 119, App
	17	18.4	61.3	65608	9	US-09-954-531-180	Sequence 180, App
	18	18.2	60.7	49507	13	US-10-087-192-1087	Sequence 1087, Ap
c	19	18	60.0	5544	13	US-10-221-613-404	Sequence 404, App
c	20	18	60.0	5544	15	US-10-240-485-174	Sequence 174, App
	21	18	60.0	237961	17	US-10-433-580-2	Sequence 2, Appli
c	22	17.8	59.3	517	13	US-10-027-632-62131	Sequence 62131, A
c	23	17.8	59.3	517	16	US-10-027-632-62131	Sequence 62131, A
c	24	17.8	59.3	524	13	US-10-027-632-297953	Sequence 297953,
c	25	17.8	59.3	524	16	US-10-027-632-297953	Sequence 297953,
	26	17.8	59.3	540	13	US-10-424-599-310	Sequence 310, App
	27	17.8	59.3	654	13	US-10-027-632-226492	Sequence 226492,
	28	17.8	59.3	654	13	US-10-027-632-226493	Sequence 226493,
	29	17.8	59.3	654	16	US-10-027-632-226492	Sequence 226492,
	30	17.8	59.3	654	16	US-10-027-632-226493	Sequence 226493,
	31	17.8	59.3	657	17	US-10-437-963-26965	Sequence 26965, A
	32	17.8	59.3	893	15	US-10-198-846-3712	Sequence 3712, Ap
c	33	17.8	59.3	1049	15	US-10-198-846-11907	Sequence 11907, A
c	34	17.8	59.3	10597	15	US-10-057-108-10	Sequence 10, Appl
c	35	17.8	59.3	10597	17	US-10-738-886-10	Sequence 10, Appl
c	36	17.8	59.3	10599	15	US-10-057-108-11	Sequence 11, Appl
c	37	17.8	59.3	10599	17	US-10-738-886-11	Sequence 11, Appl
c	38	17.8	59.3	12482	15	US-10-057-108-7	Sequence 7, Appli
c	39	17.8	59.3	12482	17	US-10-738-886-7	Sequence 7, Appli
	40	17.8	59.3	198285	9	US-09-880-107-3814	Sequence 3814, Ap
c	41	17.8	59.3	3673778	15	US-10-312-841-1	Sequence 1, Appli
c	42	17.6	58.7	291	13	US-10-027-632-53539	Sequence 53539, A
c	43	17.6	58.7	291	16	US-10-027-632-53539	Sequence 53539, A
	44	17.6	58.7	440	16	US-10-291-265-614	Sequence 614, App
	45	17.6	58.7	478	16	US-10-291-265-142	Sequence 142, App

ALIGNMENTS

RESULT 1
US-10-312-841-2/c
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (379615)
US-10-312-841-2

Query Match 64.7%; Score 19.4; DB 15; Length 3673778;
Best Local Similarity 79.3%; Pred. No. 6.2e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy          2 ACACTTAATTTGACCGAGTAACATTCGCC 30
              |||| || || ||||| |||||| ||||
Db    2679336 ACACAAAAATTAACCGAATAACATACGCC 2679308
```

RESULT 2
US-10-398-221-9/c
; Sequence 9, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik

```
;  APPLICANT: GLASER, Philippe
;  TITLE OF INVENTION: Listeria innocua, genome and applications
;  FILE REFERENCE: 344 702 - US
;  CURRENT APPLICATION NUMBER: US/10/398,221
;  CURRENT FILING DATE:  2003-03-27
;  PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: FR 00/12 697
;  PRIOR FILING DATE: 2000-10-04
;  NUMBER OF SEQ ID NOS: 4025
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
;   LENGTH: 684707
;   TYPE: DNA
;   ORGANISM: Listeria innocua
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(end)
;   OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-9

  Query Match             63.3%;  Score 19;  DB 16;  Length 684707;
  Best Local Similarity   81.5%;  Pred. No. 8.3e+02;
  Matches   22;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACACTTAATTTGACCGAGTAACATTCG 28
              || |||||||| ||| | || ||||||
Db       7378 ACGCTTAATTTTACCCATTACCATTCG 7352


RESULT 3
US-10-398-221-2058/c
; Sequence 2058, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
;  APPLICANT: KUNST, Frederik
;  APPLICANT: GLASER, Philippe
;  TITLE OF INVENTION: Listeria innocua, genome and applications
;  FILE REFERENCE: 344 702 - US
;  CURRENT APPLICATION NUMBER: US/10/398,221
;  CURRENT FILING DATE:  2003-03-27
;  PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: FR 00/12 697
;  PRIOR FILING DATE: 2000-10-04
;  NUMBER OF SEQ ID NOS: 4025
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2058
;   LENGTH: 3011208
;   TYPE: DNA
;   ORGANISM: Listeria innocua
US-10-398-221-2058

  Query Match             63.3%;  Score 19;  DB 16;  Length 3011208;
  Best Local Similarity   81.5%;  Pred. No. 8.5e+02;
  Matches   22;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACACTTAATTTGACCGAGTAACATTCG 28
              || |||||||| ||| | || ||||||
Db    2444961 ACGCTTAATTTTACCCATTACCATTCG 2444935


RESULT 4
US-10-085-783A-42095
; Sequence 42095, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
;  APPLICANT: ChondroGene Inc.
;  APPLICANT:  Liew, C.C.
;  TITLE OF INVENTION: Compositions and Methods Relatiing to Osteoarthritis
;  FILE REFERENCE: 4231/2002
;  CURRENT APPLICATION NUMBER: US/10/085,783A
;  CURRENT FILING DATE:  2002-02-28
;  PRIOR APPLICATION NUMBER: US 60/305,340
;  PRIOR FILING DATE: 2001-07-13
;  PRIOR APPLICATION NUMBER: US 60/275,017
;  PRIOR FILING DATE: 2001-03-12
;  PRIOR APPLICATION NUMBER: US 60/271,955
;  PRIOR FILING DATE: 2001-02-28
;  NUMBER OF SEQ ID NOS: 58994
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 42095
;   LENGTH: 180
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4)..(4)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6)..(6)
;   OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-42095

  Query Match             62.0%;  Score 18.6;  DB 13;  Length 180;
  Best Local Similarity   84.0%;  Pred. No. 2.2e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 ACACTTAATTTGACCGAGTAACATT 26
              | |||||||||| |||||| | ||
Db         20 AAACTTAATTTGAGCGAGTACCTTT 44


RESULT 5
US-10-242-535A-42095
; Sequence 42095, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
;  APPLICANT: ChondroGene Inc.
;  APPLICANT:  Liew, C.C.
;  TITLE OF INVENTION: Compositions and Methods Relatiing to Osteoarthritis
;  FILE REFERENCE: 4231/2005
;  CURRENT APPLICATION NUMBER: US/10/242,535A
;  CURRENT FILING DATE:  2002-09-12
;  PRIOR APPLICATION NUMBER: US 10/085,783
;  PRIOR FILING DATE: 2002-02-28
;  PRIOR APPLICATION NUMBER: US 60/305,340
;  PRIOR FILING DATE: 2001-07-13
;  PRIOR APPLICATION NUMBER: US 60/275,017
;  PRIOR FILING DATE: 2001-03-12
;  PRIOR APPLICATION NUMBER: US 60/271,955
;  PRIOR FILING DATE: 2001-02-28
;  NUMBER OF SEQ ID NOS: 58994
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 42095
;   LENGTH: 180
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4)..(4)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6)..(6)
;   OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-42095

  Query Match             62.0%;  Score 18.6;  DB 16;  Length 180;
  Best Local Similarity   84.0%;  Pred. No. 2.2e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 ACACTTAATTTGACCGAGTAACATT 26
              | |||||||||| |||||| | ||
Db         20 AAACTTAATTTGAGCGAGTACCTTT 44
```

```
RESULT 6
US-10-027-632-217354/c
; Sequence 217354, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 217354
;   LENGTH: 612
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-217354

  Query Match             62.0%;  Score 18.6;  DB 13;  Length 612;
  Best Local Similarity   84.0%;  Pred. No. 2.9e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          3 CACTTAATTTGACCGAGTAACATTC 27
              || |||||||||| || ||||| ||
Db        381 CATTTAATTTGACAGAATAACACTC 357


RESULT 7
US-10-027-632-217355/c
; Sequence 217355, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 217355
;   LENGTH: 612
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-217355

  Query Match             62.0%;  Score 18.6;  DB 13;  Length 612;
  Best Local Similarity   84.0%;  Pred. No. 2.9e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          3 CACTTAATTTGACCGAGTAACATTC 27
              || |||||||||| || ||||| ||
Db        381 CATTTAATTTGACAGAATAACACTC 357


RESULT 8
US-10-027-632-217354/c
; Sequence 217354, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 217354
;   LENGTH: 612
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-217354

  Query Match             62.0%;  Score 18.6;  DB 16;  Length 612;
  Best Local Similarity   84.0%;  Pred. No. 2.9e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          3 CACTTAATTTGACCGAGTAACATTC 27
              || |||||||||| || ||||| ||
Db        381 CATTTAATTTGACAGAATAACACTC 357


RESULT 9
US-10-027-632-217355/c
; Sequence 217355, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
```

```
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 217355
;   LENGTH: 612
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-217355

  Query Match             62.0%;  Score 18.6;  DB 16;  Length 612;
  Best Local Similarity   84.0%;  Pred. No. 2.9e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          3 CACTTAATTTGACCGAGTAACATTC 27
              || |||||||||| || ||||| ||
Db        381 CATTTAATTTGACAGAATAACACTC 357


RESULT 10
US-10-350-923B-5
; Sequence 5, Application US/10350923B
; Publication No. US20040002084A1
; GENERAL INFORMATION:
;  APPLICANT: Dotson, Stanton B.
;  APPLICANT:  Ma, Xiao Jun
;  TITLE OF INVENTION: Nucleic acid, polypeptides, vectors, and cells derived from activ
;  TITLE OF INVENTION:  eosinophil cells
;  FILE REFERENCE: S03157-01
;  CURRENT APPLICATION NUMBER: US/10/350,923B
;  CURRENT FILING DATE:  2003-01-24
;  PRIOR APPLICATION NUMBER: US/ 09/454,280
;  PRIOR FILING DATE: 1999-12-03
;  PRIOR APPLICATION NUMBER: PCT/US99/28773
;  PRIOR FILING DATE: 1999-06-12
;  PRIOR APPLICATION NUMBER: US 60/111,006
;  PRIOR FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 79
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
;   LENGTH: 868
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (34)..(37)
;   OTHER INFORMATION: n is selected from the group consisting of a, c, g, and t
US-10-350-923B-5

  Query Match             62.0%;  Score 18.6;  DB 16;  Length 868;
  Best Local Similarity   84.0%;  Pred. No. 3.1e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 ACACTTAATTTGACCGAGTAACATT 26
              | |||||||||||| |||||| | ||
Db        694 AAACTTAATTTGAGCGAGTACCTTT 718


RESULT 11
US-10-198-846-9689
; Sequence 9689, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Xu, Yongyao
;  APPLICANT:  Wang, Youzhen
;  APPLICANT:  Steinmann, Kathleen
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
;  TITLE OF INVENTION:  FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI-049
;  CURRENT APPLICATION NUMBER: US/10/198,846
;  CURRENT FILING DATE:  2002-07-18
;  PRIOR APPLICATION NUMBER: 60/306,220
;  PRIOR FILING DATE: 2001-07-18
;  NUMBER OF SEQ ID NOS: 14084
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9689
;   LENGTH: 1033
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 1028, 1029, 1030, 1031,
;   LOCATION: 1032, 1033
;   OTHER INFORMATION: n = A,T,C or G
US-10-198-846-9689

  Query Match             62.0%;  Score 18.6;  DB 15;  Length 1033;
  Best Local Similarity   84.0%;  Pred. No. 3.2e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 ACACTTAATTTGACCGAGTAACATT 26
              | |||||||||||| |||||| | ||
Db        763 AAACTTAATTTGAGCGAGTACCTTT 787


RESULT 12
US-10-311-455-1088/c
; Sequence 1088, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1088
;   LENGTH: 7500
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1088

  Query Match             61.3%;  Score 18.4;  DB 15;  Length 7500;
  Best Local Similarity   78.6%;  Pred. No. 6.1e+02;
  Matches   22;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 CACTTAATTTGACCGAGTAACATTCGCC 30
              |||||||   ||||||   |||||||||
Db       4616 CACTTAAAACGACCGAACTACATTCGCC 4589


RESULT 13
US-10-429-873A-3
; Sequence 3, Application US/10429873A
; Publication No. US20040014193A1
; GENERAL INFORMATION:
;  APPLICANT: ABUTHREIDEH, Jane et al.
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
```

```
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001364
;  CURRENT APPLICATION NUMBER: US/10/429,873A
;  CURRENT FILING DATE:  2003-05-06
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 33112
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-429-873A-3

  Query Match             61.3%;  Score 18.4;  DB 16;  Length 33112;
  Best Local Similarity   78.6%;  Pred. No. 8.4e+02;
  Matches   22;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          2 ACACTTAATTTGACCGAGTAACATTCGC 29
              || |||  | ||||| ||||||||||  |
Db      29043 ACCCTTTTTCTGACCCAGTAACATTCAC 29070


RESULT 14
US-10-199-199-14
; Sequence 14, Application US/10199199
; Publication No. US20040014047A1
; GENERAL INFORMATION:
;  APPLICANT: Lex M. Cowsert
;  APPLICANT:  Kenneth W. Dobie
;  TITLE OF INVENTION: ANTISENSE MODULATION OF LIM DOMAIN KINASE 1 EXPRESSION
;  FILE REFERENCE: RTS-0375
;  CURRENT APPLICATION NUMBER: US/10/199,199
;  CURRENT FILING DATE:  2002-07-18
;  NUMBER OF SEQ ID NOS: 148
; SEQ ID NO 14
;   LENGTH: 39001
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-199-199-14

  Query Match             61.3%;  Score 18.4;  DB 16;  Length 39001;
  Best Local Similarity   78.6%;  Pred. No. 8.7e+02;
  Matches   22;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          2 ACACTTAATTTGACCGAGTAACATTCGC 29
              || |||  | ||||| ||||||||||  |
Db      36190 ACCCTTTTTCTGACCCAGTAACATTCAC 36217


RESULT 15
US-09-962-436-292
; Sequence 292, Application US/09962436
; Patent No. US20020081301A1
; GENERAL INFORMATION:
;  APPLICANT: Soppet, Daniel
;  TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-75
;  CURRENT APPLICATION NUMBER: US/09/962,436
;  CURRENT FILING DATE:  2001-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,082
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/234,924
;  PRIOR FILING DATE: 2000-09-25
;  NUMBER OF SEQ ID NOS: 568
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 292
;   LENGTH: 65608
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: n=a,t,g or c
US-09-962-436-292

  Query Match             61.3%;  Score 18.4;  DB 9;  Length 65608;
  Best Local Similarity   78.6%;  Pred. No. 9.7e+02;
  Matches   22;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          2 ACACTTAATTTGACCGAGTAACATTCGC 29
              || |||  | ||||| ||||||||||  |
Db      36260 ACCCTTTTTCTGACCCAGTAACATTCAC 36287


Search completed: July 20, 2004, 01:48:38
Job time : 173.721 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 00:17:50 ; Search time 1374.59 Seconds
(without alignments)
651.733 Million cell updates/sec

Title: US-09-831-272-2
Perfect score: 30
Sequence: 1 cacacttaatttgaccgagtaacattcgcc 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	22.2	74.0	660	14	CF349641	CF349641 fc2569.e
2	20.6	68.7	500	14	CD581429	CD581429 RK012A4H0
3	20.2	67.3	602	12	BJ521358	BJ521358 BJ521358
c 4	20.2	67.3	682	12	BJ022103	BJ022103 BJ022103
c 5	20.2	67.3	685	12	BJ533932	BJ533932 BJ533932
c 6	20.2	67.3	740	12	BJ015002	BJ015002 BJ015002
c 7	20	66.7	982	10	BF385998	BF385998 602046177
c 8	19.6	65.3	253	9	AV084908	AV084908 AV084908
c 9	19.6	65.3	291	9	AV163173	AV163173 AV163173
c 10	19.6	65.3	556	9	AI909003	AI909003 QV-BT197-
c 11	19.6	65.3	606	10	BG080964	BG080964 H3059E08-
c 12	19.6	65.3	759	29	BX187708	BX187708 Danio rer
13	19.6	65.3	980	28	CC448991	CC448991 ZMMBBc032
c 14	19.4	64.7	389	14	CD981231	CD981231 QAI21g01.
15	19.4	64.7	494	28	AQ627538	AQ627538 CITBI-E1-
c 16	19.4	64.7	526	9	AA552611	AA552611 nk12b09.s
c 17	19.4	64.7	619	10	BE539667	BE539667 601059471
c 18	19.4	64.7	787	28	CC392698	CC392698 PUHBA16TB
19	19.4	64.7	803	28	BZ994426	BZ994426 PUGJB80TB
20	19.4	64.7	823	28	CC392700	CC392700 PUHBA16TD
c 21	19.4	64.7	827	28	BZ994430	BZ994430 PUGJB80TD
c 22	19.4	64.7	908	28	BZ722043	BZ722043 PUCDK56TD
c 23	19.2	64.0	306	10	BB208476	BB208476 BB208476
24	19.2	64.0	406	14	CD606005	CD606005 RZ147A1A1
c 25	19.2	64.0	425	28	AQ023481	AQ023481 CpGR0015B
26	19.2	64.0	548	28	AQ579759	AQ579759 T135251b
c 27	19.2	64.0	574	28	AQ571694	AQ571694 HS_2119_A
28	19.2	64.0	608	10	AW420732	AW420732 fj85h04.x
29	19.2	64.0	1201	13	BX420783	BX420783 BX420783
30	19	63.3	119	12	BM035642	BM035642 fu18c02.y
31	19	63.3	321	14	CD598710	CD598710 RK136A1D0
32	19	63.3	374	14	CD283854	CD283854 G39167.93
33	19	63.3	405	14	CB359807	CB359807 ZF001-P00
34	19	63.3	424	12	BM279675	BM279675 zah3054 Z
35	19	63.3	494	10	BF938422	BF938422 fm76e06.y
36	19	63.3	502	12	BG307726	BG307726 fm52b07.y
37	19	63.3	505	12	BG305109	BG305109 fl94g02.y
38	19	63.3	510	13	BQ261533	BQ261533 fz71g10.y
39	19	63.3	528	12	BI672771	BI672771 ft44c06.y
40	19	63.3	540	12	BM530029	BM530029 fw95h11.y
41	19	63.3	556	12	BM530969	BM530969 fw99a04.y
42	19	63.3	614	12	BG307193	BG307193 fm13c02.y
43	19	63.3	614	12	BM775766	BM775766 fy41a10.y
44	19	63.3	649	13	BU710008	BU710008 fv23a09.y
45	19	63.3	654	12	BM860238	BM860238 fy35g11.y

ALIGNMENTS

RESULT 1
CF349641
LOCUS CF349641 660 bp mRNA linear EST 18-AUG-2003
DEFINITION fc2569.e Rose Petals (Fragrant Cloud) Lambda Zap Express Library Rosa hybrid cultivar cDNA clone fc2569.e 5', mRNA sequence.
ACCESSION CF349641
VERSION CF349641.1 GI:33832991
KEYWORDS EST.
SOURCE Rosa hybrid cultivar
ORGANISM Rosa hybrid cultivar
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Rosales; Rosaceae; Rosoideae; Rosa.
REFERENCE 1 (bases 1 to 660)
AUTHORS Guterman,I., Shalit,M., Menda,N., Piestun,D., Dafny-Yelin,M., Shalev,G., Bar,E., Davydov,O., Ovadis,M., Emanuel,M., Wang,J., Adam,Z., Pichersky,E., Lewinsohn,E., Zamir,D., Vainstein,A. and Weiss,D.
TITLE Rose scent: genomics approach to discovering novel floral fragrance-related genes
JOURNAL Plant Cell 14 (10), 2325-2338 (2002)
MEDLINE 22256146
PUBMED 12368489
COMMENT Contact: Naama Menda
Petal Genomics
Faculty of Agricultural, Food and Environmental Quality Sciences, The Hebrew University of Jerusalem

```
           P.O. Box 12, Rehovot, 76100, Israel
           Tel: 972 8 9489 683
           Fax: 972 8 9468 265
           Email: shaham@agri.huji.ac.il
           Seq primer: T3 forward.
FEATURES             Location/Qualifiers
     source          1. .660
                     /organism="Rosa hybrid cultivar"
                     /mol_type="mRNA"
                     /strain="Fragrant Cloud"
                     /db_xref="taxon:128735"
                     /clone="fc2569.e"
                     /tissue_type="Petals"
                     /dev_stage="Young open flower at stage four"
                     /clone_lib="Rose Petals (Fragrant Cloud) Lambda Zap
                     Express Library"
                     /note="Vector: pBKCMV; Site_1: EcoRI; Site_2: XhoI"
ORIGIN

  Query Match             74.0%;  Score 22.2;  DB 14;  Length 660;
  Best Local Similarity   88.9%;  Pred. No. 26;
  Matches   24;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 CACTTAATTTGACCGAGTAACATTCGC 29
              ||||||||| | |||||||||||| ||
Db        441 CACTTAATTAGGCCGAGTAACATTGGC 467


RESULT 2
CD581429
LOCUS       CD581429                500 bp    mRNA    linear   EST 16-JUN-2003
DEFINITION  RK012A4H08.T3 Zebrafish Kidney Marrow cDNA library Danio rerio cDNA
            clone RK012A4H08 5', mRNA sequence.
ACCESSION   CD581429
VERSION     CD581429.1  GI:31758092
KEYWORDS    EST.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Song,H.D., Wu,X.Y., Sun,X.J., Zhou,Y., Liu,T.X., Zhang,G.W.,
            Sheng,Y., Chen,Y., Ruan,Z., Jiang,C.L., Fan,H.Y., Look,A.T.,
            Zon,L.I. and Chen,Z.
  TITLE     Gene Expression Profiling in the Zebrafish Kidney Marrow Tissue
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Chen Z.
            State Key Lab for Medical Genomics
            Shanghai Institute of Hematology, Ruijin Hospital Affiliated to
            Shanghai Second Medical University
            197 Rui Jin Road II, Shanghai 200025, P. R. China
            Tel: 86-21-64740490
            Fax: 86-21-64743206
            Email: zchen@stn.sh.cn
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="RK012A4H08"
                     /dev_stage="mature"
                     /clone_lib="Zebrafish Kidney Marrow cDNA library"
                     /note="Organ: kidney; Vector: pBS-CMV; Site_1: XhoI;
                     Site_2: EcoRI; Total RNA was extracted from the kidney
                     tissues of mature zebrafish. The poly (A)+ RNA fraction
                     was separated from total RNA by oligo (dT) cellulose
                     chromatography. Library was initially constructed in the
                     lambdaZAP Express vector (Stratagene) and in vivo excised
                     into pBS-CMV vector."
ORIGIN

  Query Match             68.7%;  Score 20.6;  DB 14;  Length 500;
  Best Local Similarity   85.2%;  Pred. No. 1.3e+02;
  Matches   23;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          3 CACTTAATTTGACCGAGTAACATTCGC 29
              |||||  |||||||||||| || |||||
Db         22 CACTTCTTTTGACCGAGTCACCTTCGC 48


RESULT 3
BJ521358
LOCUS       BJ521358                602 bp    mRNA    linear   EST 09-AUG-2002
DEFINITION  BJ521358 MF01SSB cDNA Oryzias latipes cDNA clone MF01SSB018H19 5',
            mRNA sequence.
ACCESSION   BJ521358
VERSION     BJ521358.1  GI:22180170
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 602)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .602
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSB018H19"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSB cDNA"
ORIGIN

  Query Match             67.3%;  Score 20.2;  DB 12;  Length 602;
  Best Local Similarity   88.0%;  Pred. No. 2e+02;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACAT 25
              | |||||||||||||||||| | |||
Db        228 CTCACTTAATTTGACCGAGAAGCAT 252


RESULT 4
BJ022103/c
LOCUS       BJ022103                682 bp    mRNA    linear   EST 05-DEC-2001
DEFINITION  BJ022103 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA118F02 3',
            mRNA sequence.
ACCESSION   BJ022103
VERSION     BJ022103.1  GI:17371006
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 682)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
```

```
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .682
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSA118F02"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSA cDNA"
ORIGIN

  Query Match           67.3%;  Score 20.2;  DB 12;  Length 682;
  Best Local Similarity   88.0%;  Pred. No. 2e+02;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACAT 25
              | ||||||||||||||||| | |||
Db        440 CTCACTTAATTTGACCGAGAAGCAT 416


RESULT 5
BJ533932/c
LOCUS       BJ533932                685 bp    mRNA    linear   EST 09-AUG-2002
DEFINITION  BJ533932 MF01SSB cDNA Oryzias latipes cDNA clone MF01SSB018H19 3',
            mRNA sequence.
ACCESSION   BJ533932
VERSION     BJ533932.1  GI:22192744
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 685)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .685
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSB018H19"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSB cDNA"
ORIGIN

  Query Match           67.3%;  Score 20.2;  DB 12;  Length 685;
  Best Local Similarity   88.0%;  Pred. No. 2e+02;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACAT 25
              | ||||||||||||||||| | |||
Db        461 CTCACTTAATTTGACCGAGAAGCAT 437


RESULT 6
BJ015002/c
LOCUS       BJ015002                740 bp    mRNA    linear   EST 07-DEC-2001
DEFINITION  BJ015002 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA017F10 3',
            mRNA sequence.
ACCESSION   BJ015002
VERSION     BJ015002.1  GI:17406329
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 740)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .740
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSA017F10"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSA cDNA"
ORIGIN

  Query Match           67.3%;  Score 20.2;  DB 12;  Length 740;
  Best Local Similarity   88.0%;  Pred. No. 2e+02;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACAT 25
              | ||||||||||||||||| | |||
Db        441 CTCACTTAATTTGACCGAGAAGCAT 417


RESULT 7
BF385998/c
LOCUS       BF385998                982 bp    mRNA    linear   EST 27-NOV-2000
DEFINITION  602046177F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:4195675 5',
            mRNA sequence.
ACCESSION   BF385998
VERSION     BF385998.1  GI:11367303
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 982)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
```

```
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM9531  row: i  column: 20
           High quality sequence stop: 667.
FEATURES             Location/Qualifiers
     source          1. .982
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4195675"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Li9"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

  Query Match             66.7%;  Score 20;  DB 10;  Length 982;
  Best Local Similarity   82.1%;  Pred. No. 2.7e+02;
  Matches   23;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 CACTTAATTTGACCGAGTAACATTCGCC 30
              |||||||| |  || |||||| ||||||
Db        822 CACTTAATCTTTCCCAGTAACGTTCGCC 795


RESULT 8
AV084908/c
LOCUS       AV084908                253 bp    mRNA    linear   EST 25-JUN-1999
DEFINITION  AV084908 Mus musculus tongue C57BL/6J adult Mus musculus cDNA clone
            2310009B15, mRNA sequence.
ACCESSION   AV084908
VERSION     AV084908.1  GI:5216356
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 253)
  AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
            Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
            Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
            Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
            Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
            Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
            Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .253
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="2310009B15"
                     /sex="male"
                     /tissue_type="tongue"
                     /dev_stage="adult"
                     /clone_lib="Mus musculus tongue C57BL/6J adult"
ORIGIN

  Query Match             65.3%;  Score 19.6;  DB 9;  Length 253;
  Best Local Similarity   84.6%;  Pred. No. 3e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATT 26
              |||||||||||||| || ||| | |
Db        225 CACACTTAATTTGACAGACTAAAAAT 200


RESULT 9
AV163173/c
LOCUS       AV163173                291 bp    mRNA    linear   EST 08-JUL-1999
DEFINITION  AV163173 Mus musculus head C57BL/6J 13-day embryo Mus musculus cDNA
            clone 3110007I01, mRNA sequence.
ACCESSION   AV163173
VERSION     AV163173.1  GI:5429149
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
            Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
            Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
            Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
            Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
            Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
            Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="3110007I01"
                     /sex="mixed"
                     /tissue_type="head"
                     /dev_stage="13-day embryo"
                     /clone_lib="Mus musculus head C57BL/6J 13-day embryo"
ORIGIN

  Query Match             65.3%;  Score 19.6;  DB 9;  Length 291;
  Best Local Similarity   84.6%;  Pred. No. 3.1e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATT 26
              |||||||||||||| || ||| | |
Db        265 CACACTTAATTTGACAGACTAAAAAT 240
```

```
RESULT 10
AI909003/c
LOCUS       AI909003                556 bp    mRNA    linear   EST 30-MAR-2000
DEFINITION  QV-BT197-050499-008 BT197 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AI909003
VERSION     AI909003.1  GI:6499683
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/seq/gethtml.pl?t1=QV&t2=QV-BT197-008.html
            &t3=050499&t4=1)
            Seq primer: puc 18 forward.
FEATURES             Location/Qualifiers
     source          1. .556
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /sex="female"
                     /dev_stage="Adult"
                     /clone_lib="BT197"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No.
                     196,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
ORIGIN

  Query Match           65.3%;  Score 19.6;  DB 9;  Length 556;
  Best Local Similarity   84.6%;  Pred. No. 3.6e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATT 26
              || |||||||||| |||| | |||||
Db        263 CAGACTTAATTTGCCCGAATTACATT 238


RESULT 11
BG080964/c
LOCUS       BG080964                606 bp    mRNA    linear   EST 17-DEC-2003
DEFINITION  H3059E08-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
            H3059E08 5', mRNA sequence.
ACCESSION   BG080964
VERSION     BG080964.2  GI:40015143
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 606)
  AUTHORS   Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X.,
            Grahovac,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H.,
            Wood,W.H. III, Becker,K.G. and Ko,M.S.H.
  TITLE     Genome-wide expression profiling of mid-gestation placenta and
            embryo using a 15,000 mouse developmental cDNA microarray
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
  MEDLINE   20381348
   PUBMED   10922068
COMMENT     On Jan 26, 2001 this sequence version replaced gi:12563532.
            Other_ESTs: H3059E08-3
            Contact: George J. Kargul
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            This clone set has been freely distributed to the community. Please
            visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
            Plate: H3059  row: E  column: 08
            Seq primer: -21M13 Reverse
            High quality sequence stop: 606
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .606
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:H3059E08-5"
                     /db_xref="taxon:10090"
                     /clone="H3059E08"
                     /sex="Clones arrayed from a variety of cDNA libraries"
                     /dev_stage="Clones arrayed from a variety of cDNA
                     libraries"
                     /lab_host="DH10B"
                     /clone_lib="NIA Mouse 15K cDNA Clone Set"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     clone is among a rearrayed set of 15,247 clones from 11
                     embryo  cDNA libraries (including preimplantation stage
                     embryos from  unfertilized egg to blastocyst, embryonic
                     part of E7.5 embryos,  extraembryonic part of E7.5
                     embryos, and E12.5 female mesonephros/gonad)  and one
                     newborn ovary cDNA library. Average insert size 1.5 kb.
                     All  source libraries are cloned unidirectionally with
                     Oligo(dT)-Not primers.  References include: (1)
                     Genome-wide expression profiling of mid-gestation
                     placenta and embryo using a 15,000 mouse developmental
                     cDNA microarray,  2000, Proc. Natl. Acad. Sci. U S A, 97:
                     9127-9132; (2) Large-scale cDNA  analysis reveals phased
                     gene expression patterns during preimplantation  mouse
                     develolpment, 2000, Development, 127: 1737-1749; (3)
                     Genome-wide  mapping of unselected transcripts from
                     extraembryonic tissue of 7.5-day  mouse embryos reveals
                     enrichment in the t-complex and under-representation  on
                     the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."
ORIGIN

  Query Match           65.3%;  Score 19.6;  DB 10;  Length 606;
  Best Local Similarity   84.6%;  Pred. No. 3.6e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          3 CACTTAATTTGACCGAGTAACATTCG 28
              ||||||| || |||||||| |||| |
Db         37 CACTTAAGTTTACCGAGTACCATTTG 12


RESULT 12
BX187708/c
LOCUS       BX187708                759 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-194P17, genomic survey sequence.
ACCESSION   BX187708
VERSION     BX187708.1  GI:28019594
```

```
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 759)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 194P17. 194P17
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .759
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-194P17"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             65.3%;  Score 19.6;  DB 29;  Length 759;
  Best Local Similarity   84.6%;  Pred. No. 3.8e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATT 26
              ||| || ||| | ||||||||||||||
Db        505 CACTCTAAATATAACCGAGTAACATT 480


RESULT 13
CC448991
LOCUS       CC448991                980 bp    DNA     linear   GSS 22-MAY-2003
DEFINITION  ZMMBBc0322H05f ZMMBBc Zea mays subsp. mays genomic clone
            ZMMBBc0322H05 5', genomic survey sequence.
ACCESSION   CC448991
VERSION     CC448991.1  GI:30996554
KEYWORDS    GSS.
SOURCE      Zea mays subsp. mays (maize)
  ORGANISM  Zea mays subsp. mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 980)
  AUTHORS   Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
            Rouzard,K., Fuks,G., Yu,Y., Wing,R. and Messing,J.
  TITLE     Sequencing of the maize genome at PGIR (2003b)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Bharti,A.K.
            Dr.Joachim Messing's lab
            The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
            University
            190 Frelinghuysen Road, Piscataway, NJ 08854, USA
            Tel: 732 445 3801
            Fax: 732 445 5735
            Email: bharti@waksman.rutgers.edu
            Seq primer: T7
            Class: BAC ends
            High quality sequence start: 3.
FEATURES             Location/Qualifiers
     source          1. .980
                     /organism="Zea mays subsp. mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /sub_species="mays"
                     /db_xref="taxon:4578"
                     /clone="ZMMBBc0322H05"
                     /lab_host="E. coli DH10B"
                     /clone_lib="ZMMBBc"
                     /note="Vector: pTARBAC1.3; Site_1: BamHI; Site_2: BamHI"
ORIGIN

  Query Match             65.3%;  Score 19.6;  DB 28;  Length 980;
  Best Local Similarity   84.6%;  Pred. No. 4e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATT 26
              ||||||||||||||  |||||  ||
Db        808 CACACTTAATTTGACAAAGTAAACTT 833


RESULT 14
CD981231/c
LOCUS       CD981231                389 bp    mRNA    linear   EST 16-JUL-2003
DEFINITION  QAI21g01.yg QAI Zea mays cDNA clone QAI21g01, mRNA sequence.
ACCESSION   CD981231
VERSION     CD981231.1  GI:32841550
KEYWORDS    EST.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 389)
  AUTHORS   Genoplante.
  TITLE     Genoplante, a major partnership french program in plant genomics
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Genoplante
            Genoplante
            93, rue Henri Rochefort 91025 EVRY CEDEX France
            Tel: 33 1 69 47 54 00
            Fax: 33 1 69 47 54 10
            This sequence has been generated in the framework of the french
            plant genomics programme 'Genoplante' (http://www.genoplante.com
            and http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .389
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /cultivar="F2"
                     /db_xref="taxon:4577"
                     /clone="QAI21g01"
                     /tissue_type="ear leaf"
                     /clone_lib="QAI"
ORIGIN

  Query Match             64.7%;  Score 19.4;  DB 14;  Length 389;
  Best Local Similarity   79.3%;  Pred. No. 4.1e+02;
  Matches   23;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          2 ACACTTAATTTGACCGAGTAACATTCGCC 30
              |||  |  |||||||||||||| ||||| ||
Db        183 ACAAATGTTTTGACCGAGTATCATTCACC 155


RESULT 15
AQ627538
LOCUS       AQ627538                494 bp    DNA     linear   GSS 16-JUN-1999
DEFINITION  CITBI-E1-2653P15.TF CITBI-E1 Homo sapiens genomic clone 2653P15,
            genomic survey sequence.
ACCESSION   AQ627538
VERSION     AQ627538.1  GI:5089930
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
```

```
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CITBI-E1-2653P15.TR
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .494
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /clone="2653P15"
                      /sex="male"
                      /cell_type="sperm"
                      /clone_lib="CITBI-E1"
                      /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                      CalTech Human BAC Library D"
ORIGIN

  Query Match          64.7%;  Score 19.4;  DB 28;  Length 494;
  Best Local Similarity   79.3%;  Pred. No. 4.3e+02;
  Matches   23;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATTCGC 29
              |||| |||| || ||||||| ||||| |
Db        303 CACATGTAATGTGTCCGAGTAGCATTCCC 331


Search completed: July 20, 2004, 01:42:45
Job time : 1377.59 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 00:17:50 ; Search time 1420.41 Seconds
(without alignments)
651.733 Million cell updates/sec

Title: US-09-831-272-11
Perfect score: 31
Sequence: 1 tacaattcaaacattgttcaaacaaggaacc 31

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	23.2	74.8	844	28	BZ392622	BZ392622 EINBH33TF
	2	22	71.0	568	28	AZ109637	AZ109637 RPCI-23-1
c	3	22	71.0	706	28	BZ676143	BZ676143 PUBBB31TD
c	4	22	71.0	842	28	BZ669683	BZ669683 PUBAV94TD
c	5	21.8	70.3	489	29	CPA561291	AJ561291 Cryptospo
	6	21.6	69.7	572	10	BG044802	BG044802 saa32g02.
	7	21.6	69.7	587	12	BQ040901	BQ040901 gd39h11.y
c	8	21.6	69.7	675	10	BE824450	BE824450 GM700023B
c	9	21.6	69.7	687	13	CA130450	CA130450 SCCCRT100
	10	21.6	69.7	689	28	AQ780940	AQ780940 HS_3138_B
c	11	21.6	69.7	734	28	AZ202685	AZ202685 SP_0059_B
c	12	21.4	69.0	283	9	AV961084	AV961084 AV961084
	13	21.4	69.0	407	12	BI161768	BI161768 RE01178.3
c	14	21.4	69.0	486	13	BW195405	BW195405 BW195405
c	15	21.4	69.0	486	13	BW209523	BW209523 BW209523
c	16	21.4	69.0	560	9	AV963578	AV963578 AV963578
c	17	21.4	69.0	561	13	BW274645	BW274645 BW274645
c	18	21.4	69.0	581	13	BW277525	BW277525 BW277525
c	19	21.4	69.0	587	9	AV992847	AV992847 AV992847
c	20	21.4	69.0	591	13	BW288634	BW288634 BW288634
c	21	21.4	69.0	612	9	AV999016	AV999016 AV999016
c	22	21.4	69.0	614	9	AV673090	AV673090 AV673090
c	23	21.4	69.0	620	13	BW243946	BW243946 BW243946
c	24	21.4	69.0	623	9	AV951238	AV951238 AV951238
c	25	21.4	69.0	627	9	AV951117	AV951117 AV951117
c	26	21.4	69.0	638	13	BW291174	BW291174 BW291174
c	27	21.4	69.0	665	13	BW213642	BW213642 BW213642
c	28	21.4	69.0	699	29	BX140539	BX140539 Danio rer
	29	21.4	69.0	703	12	BI573756	BI573756 RH09280.3
c	30	21.4	69.0	737	13	BW272276	BW272276 BW272276
	31	21.4	69.0	804	29	CNS009QL	AL054414 Drosophil
	32	21.4	69.0	891	28	AZ538099	AZ538099 ENTFG90TF
	33	21.4	69.0	944	28	BH151913	BH151913 ENTQD62TR
c	34	21.2	68.4	245	14	CA246963	CA246963 SCSGFL5C0
c	35	21.2	68.4	245	14	CA248260	CA248260 SCCCFL509
c	36	21.2	68.4	304	13	CA149735	CA149735 SCCCRZ200
c	37	21.2	68.4	362	14	CF090483	CF090483 QHM4P15.y
c	38	21.2	68.4	507	14	CA244864	CA244864 SCSFFL408
c	39	21.2	68.4	527	14	CA189166	CA189166 SCCCLB1C0
c	40	21.2	68.4	530	13	CA115205	CA115205 SCSBLB103
	41	21.2	68.4	539	14	CA223116	CA223116 SCBGFL404
	42	21.2	68.4	584	14	CA240951	CA240951 SCUTFL307
	43	21.2	68.4	585	14	CA223520	CA223520 SCJFFL1C0
	44	21.2	68.4	665	28	BH005742	BH005742 BMBAC10A2
	45	21.2	68.4	673	14	CA233741	CA233741 SCSGFL4C0

ALIGNMENTS

RESULT 1
BZ392622
LOCUS BZ392622 844 bp DNA linear GSS 30-APR-2003
DEFINITION EINBH33TF EI_10_12_KB Entamoeba invadens genomic clone EINBH33, genomic survey sequence.
ACCESSION BZ392622
VERSION BZ392622.1 GI:30239159
KEYWORDS GSS.
SOURCE Entamoeba invadens
ORGANISM Entamoeba invadens
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 844)
AUTHORS Wang,Z., Samuelson,J., Clark,C.G., Eichinger,D., Paul,J., van Dellen,K., Hall,N., Anderson,I. and Loftus,B.
TITLE Gene discovery in the Entamoeba invadens genome
JOURNAL Mol. Biochem. Parasitol. 129 (1), 23-31 (2003)
MEDLINE 22684048
PUBMED 12798503
COMMENT Other_GSSs: EINBH33TR
Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
DNA was provided by Daniel Eichinger

```
           Seq primer: TF
           Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Entamoeba invadens"
                     /mol_type="genomic DNA"
                     /strain="IP-1"
                     /db_xref="taxon:33085"
                     /clone="EINBH33"
                     /clone_lib="EI_10_12_KB"
                     /note="Vector: pHOS2; Site_1: BstXI; Total genomic DNA was
                     isolated from early log phase trophozoites of E. invadens
                     IP-1 using a Qiagen plant DNA extraction kit. A shotgun
                     medium-size plasmid library (average insert size of 10 -
                     12 kb) was generated by random mechanical shearing of E.
                     invadens genomic DNA, repairing the ends of DNA fragments
                     with T4 Polymerase, adding BstXI adaptors and ligating
                     into the BstXI site of a pUC-derived vector pHOS2."
ORIGIN

  Query Match             74.8%;  Score 23.2;  DB 28;  Length 844;
  Best Local Similarity   89.3%;  Pred. No. 1.7e+02;
  Matches   25;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          4 AATTCAAACATTGTTCAAACAAGGAACC 31
              |||| |||||||  ||||||||||||||
Db        716 AATTAAAACATTAGTCAAACAAGGAACC 743


RESULT 2
AZ109637
LOCUS       AZ109637                 568 bp    DNA     linear   GSS 09-MAY-2000
DEFINITION  RPCI-23-12E1.TV RPCI-23 Mus musculus genomic clone RPCI-23-12E1,
            genomic survey sequence.
ACCESSION   AZ109637
VERSION     AZ109637.1  GI:7762695
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 568)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S.,
            Akinret,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de
            Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 12  row: E  column: 1
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .568
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-12E1"
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_lib="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
ORIGIN

  Query Match             71.0%;  Score 22;  DB 28;  Length 568;
  Best Local Similarity   83.3%;  Pred. No. 4.8e+02;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGAAC 30
              |||||| |||| || || || |||||||||
Db        417 TACAATGCAAAGATGGTGCACACAAGGAAC 446


RESULT 3
BZ676143/c
LOCUS       BZ676143                 706 bp    DNA     linear   GSS 05-FEB-2003
DEFINITION  PUBBB31TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa014F14,
            genomic survey sequence.
ACCESSION   BZ676143
VERSION     BZ676143.1  GI:28227523
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 706)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .706
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa014F14"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             71.0%;  Score 22;  DB 28;  Length 706;
  Best Local Similarity   83.3%;  Pred. No. 4.8e+02;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGAAC 30
              || |||| |||| || |||||||||| ||||
Db         85 TAAAATTTAAACCTTTTTCAAACAAAGAAC 56


RESULT 4
BZ669683/c
LOCUS       BZ669683                 842 bp    DNA     linear   GSS 05-FEB-2003
DEFINITION  PUBAV94TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa013O20,
            genomic survey sequence.
ACCESSION   BZ669683
VERSION     BZ669683.1  GI:28216592
```

```
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 842)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .842
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa013O20"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             71.0%;  Score 22;  DB 28;  Length 842;
  Best Local Similarity   83.3%;  Pred. No. 4.7e+02;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGAAC 30
              || |||| |||| || |||||||||| ||||
Db         46 TAAAATTTAAACCTTTTTCAAACAAAGAAC 17


RESULT 5
CPA561291/c
LOCUS       CPA561291               489 bp    DNA     linear   GSS 11-AUG-2003
DEFINITION  Cryptosporidium parvum GSS, PAC clone pica_0001_d10, T7 end
            sequence, genomic survey sequence.
ACCESSION   AJ561291
VERSION     AJ561291.1  GI:31335740
KEYWORDS    GSS; genome survey sequence.
SOURCE      Cryptosporidium parvum
  ORGANISM  Cryptosporidium parvum
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
            Cryptosporidiidae; Cryptosporidium.
REFERENCE   1
  AUTHORS   Bankier,A.T., Spriggs,H.F., Fartmann,B., Konfortov,B.A., Madera,M.,
            Vogel,C., Teichmann,S.A., Ivens,A. and Dear,P.H.
  TITLE     Integrated mapping, chromosomal sequencing and sequence analysis of
            Cryptosporidium parvum
  JOURNAL   Genome Res. 13 (8), 1787-1799 (2003)
  MEDLINE   22784085
   PUBMED   12869580
REFERENCE   2  (bases 1 to 489)
  AUTHORS   Dear,P.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Dear P.H., PNAC Biotech Division, MRC
            Laboratory of Molecular Biology, Hills Road, Cambridge, Cambs CB2
            2QH, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Cryptosporidium parvum"
                     /mol_type="genomic DNA"
                     /serotype="Type 2"
                     /isolate="Iowa"
                     /db_xref="taxon:5807"
                     /clone="pica_0001_d10"
                     /clone_lib="pica"
                     /dev_stage="oocyst"
                     /note="T7 end sequence"
ORIGIN

  Query Match             70.3%;  Score 21.8;  DB 29;  Length 489;
  Best Local Similarity   92.0%;  Pred. No. 5.8e+02;
  Matches   23;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          5 ATTCAAACATTGTTCAAACAAGGAA 29
              ||||||||||| |||||||||| |||
Db         61 ATTCAAACATTCTTCAAACAATGAA 37


RESULT 6
BG044802
LOCUS       BG044802                572 bp    mRNA    linear   EST 28-NOV-2001
DEFINITION  saa32g02.y1 Gm-c1059 Glycine max cDNA clone GENOME SYSTEMS CLONE
            ID: Gm-c1059-700 5', mRNA sequence.
ACCESSION   BG044802
VERSION     BG044802.1  GI:12491894
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 572)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            High quality sequence stop: 421.
FEATURES             Location/Qualifiers
     source          1. .572
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1059-700"
                     /tissue_type="Whole seedling, 2 week old, etiolated"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1059"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     from 2 week old etiolated whole seedlings of PI468916.
                     Complementary DNA was synthesized from mRNA using a primer
                     consisting of a poly(dT) sequence with a XhoI restriction
                     site. EcoRI adapters were ligated to the blunt-ended cDNA
                     fragments followed by XhoI digestion. The cDNA fragments
                     were directionally cloned into the EcoRI- XhoI restriction
                     site of the pBluescript vector. The ligated cDNA fragments
                     were transformed into DH10B host cells (Gibco BRL). This
                     library was  constructed in the laboratory of Dr. Randy
                     Shoemaker at Iowa state university."
ORIGIN
```

```
  Query Match             69.7%;  Score 21.6;  DB 10;  Length 572;
  Best Local Similarity   85.7%;  Pred. No. 6.8e+02;
  Matches   24;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          4 AATTCAAACATTGTTCAAACAAGGAACC 31
              |||||| | ||||||||| ||||||| |
Db        256 AATTCAGAAATTGTTCAAGCAAGGAAAC 283


RESULT 7
BQ040901
LOCUS       BQ040901                587 bp    mRNA    linear   EST 17-APR-2002
DEFINITION  gd39h11.y1 Moss EST library PPS Physcomitrella patens cDNA clone
            PEP_SOURCE_ID:PPS101622 5' similar to TR:P93818 P93818 HYPOTHETICAL
            50.9 KD PROTEIN. ;, mRNA sequence.
ACCESSION   BQ040901
VERSION     BQ040901.1  GI:19788753
KEYWORDS    EST.
SOURCE      Physcomitrella patens
  ORGANISM  Physcomitrella patens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1  (bases 1 to 587)
  AUTHORS   Quatrano,R., Bashiardes,S., Cove,D., Cuming,A., Knight,C.,
            Clifton,S., Marra,M., Hillier,L., Pape,D., Martin,J., Wylie,T.,
            Underwood,K., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Steptoe,M., Gibbons,M., Harvey,N., Ritter,E.,
            Jackson,Y., McCann,R., Waterston,R. and Wilson,R.
  TITLE     Leeds/Wash U Moss EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: gd39h11.x1
            Contact: Ralph Quatrano
            Leeds/Wash U Moss EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Libraries were constructed by Dr. Stavros Bashiardes as part of the
            Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
            Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
            University Genome Sequencing Center For information on obtaining a
            clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
            Possible reversed clone: similarity on wrong strand
            High quality sequence stop: 332.
FEATURES             Location/Qualifiers
     source          1. .587
                     /organism="Physcomitrella patens"
                     /mol_type="mRNA"
                     /db_xref="taxon:3218"
                     /clone="PEP_SOURCE_ID:PPS101622"
                     /dev_stage="protonemata, 7day old untreated"
                     /lab_host="E.coli DH10b"
                     /clone_lib="Moss EST library PPS"
                     /note="Vector: pBluescript SK-; Site_1: XhoI; Site_2:
                     EcoRI; Library constructed by Stavros Bashiardes and
                     re-arrayed by A. Cuming & Honglin Rong. Construction of
                     the cDNA library was carried out using Statagene's 'UniZAP
                     - cDNA synthesis kit' to ligate cDNA directionally in
                     UniZAP XR vector arms. The vector is designed containing
                     the pBluescript sequence as well as the lambda DNA and
                     cDNA is cloned in the EcoRI and XhoI sites in the
                     pBluescript sequence. The vector was then packaged using
                     Gold gigapackaging extracts, propagated in XL-IBlue MRF
                     cels and amplified. The library was excied by mass
                     excision using Stratagene's Mass excision kit to infect
                     SOLR cells with phagemids and ampicillin resistant
                     transformants selected. Approximately 1,000,000 colonies
                     were grown and recovered by using Quiagen midi prep kit.2
                     micro grams of plasmid DNA were used to transform DH10b
                     cells by electroporation. Clones corresponding to abundant
                     transcripts were identified by colony hybridization and
                     eliminated from the library, be rearraying. This library
                     is non-directionally cloned."
ORIGIN

  Query Match             69.7%;  Score 21.6;  DB 12;  Length 587;
  Best Local Similarity   85.7%;  Pred. No. 6.7e+02;
  Matches   24;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          4 AATTCAAACATTGTTCAAACAAGGAACC 31
              | ||||||||||||||| |||||| || |
Db        496 ACTTCAAACATTGTTCGAACAAGTAAAC 523


RESULT 8
BE824450/c
LOCUS       BE824450                675 bp    mRNA    linear   EST 24-MAY-2001
DEFINITION  GM700023B20D7 Gm-r1070 Glycine max cDNA clone Gm-r1070-9014 3',
            mRNA sequence.
ACCESSION   BE824450
VERSION     BE824450.1  GI:10256684
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 675)
  AUTHORS   Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
            Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
  TITLE     A Functional Genomics Program for Soybean (NSF 9872565)
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: AW568213 corresponding to Gm-r1030-4141 (5')
            Contact: Vodkin, L.O., PI, A Functional Genomics Program for
            Soybean (NSF 9872565)
            Lewin, H. A., Director, Keck Center for Comparative and Functional
            Genomics
            University of Illinois
            Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
            Tel: (217) 244-6147
            Fax: (217) 333-4582
            Email: l-vodkin@uiuc.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134. For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact:clones@genomesystems.com or info@genome
            ystems.com web site:www.genomesystems.com
            Seq primer: 5'-TTTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES             Location/Qualifiers
     source          1. .675
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
                     /clone="Gm-r1070-9014"
                     /clone_lib="Gm-r1070"
                     /note="The library Gm-r1070 is a sequence-driven, reracked
                     set of 9,216  clones selected from cDNA libraries from
                     various tissues and stages  of development of soybean that
                     represent 2,639 sequences from  immature cotyledons, 1,770
                     from immature seed coats, 3,938 from  flowers, and 869
                     from young pods.  The 5' ESTs of the source clones from
                     the different libraries was  used to select singletons, or
                     a representative of each contig, which  were reracked to
                     form library Gm-r1070.  The cDNA clones of the reracked
                     Gm-r1070 library were then sequenced  at the 3' end.  The
                     contig analysis to select unique genes was performed by
                     the  laboratory of Ernest Retzel, Center for Computational
                     Genomics and  Bioinformatics, University of Minnesota,
                     http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
                     .  Reracking was performed by Genome Systems, St. Louis,
                     http://www.genomesystems.com, and 3' sequencing by the
                     Keck Center  for Comparative and Functional Genomics,
                     University of Illinois,
                     http://www.life.uiuc.edu/biotech/keck.html.  Note: The
```

corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

ORIGIN

Query Match 69.7%; Score 21.6; DB 10; Length 675;
Best Local Similarity 85.7%; Pred. No. 6.7e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy          4 AATTCAAACATTGTTCAAACAAGGAACC 31
              |||||| | ||||||||| ||||||| |
Db        433 AATTCAGAAATTGTTCAAGCAAGGAAAC 406
```

RESULT 9
CA130450/c
LOCUS CA130450 687 bp mRNA linear EST 24-SEP-2003
DEFINITION SCCCRT1002A10.g RT1 Saccharum officinarum cDNA clone SCCCRT1002A10 5', mRNA sequence.
ACCESSION CA130450
VERSION CA130450.1 GI:35013327
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Saccharum.
REFERENCE 1 (bases 1 to 687)
AUTHORS Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
TITLE The libraries that made SUCEST
JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parruda@unicamp.br
Clone distribution: clone distribution information can be found through the Brazilian Clone Collection Center (BCCC) at http://www.bcccenter.fcav.unesp.br
Plate: 002 row: A column: 10
Seq primer: T7 Promoter Primer.
FEATURES Location/Qualifiers
source 1. .687
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCCCRT1002A10"
/lab_host="DH10B"
/clone_lib="RT1"
/note="Organ: Root tips (0.3cm-long) from adult plants; Vector: pSport1; Site_1: SalI; Site_2: NotI; An unidirectional cDNA library generated from [Root tips (0.3cm-long) from adult plants]. cDNA was prepared from polyA+ mRNA using SuperScript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at http://sucest.lad.ic.unicamp.br/public"
ORIGIN

Query Match 69.7%; Score 21.6; DB 13; Length 687;
Best Local Similarity 85.7%; Pred. No. 6.7e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy          3 CAATTCAAACATTGTTCAAACAAGGAAC 30
              |||| ||| ||||||||| ||| |||||
Db        655 CAATCCAACCATTGTTCAGACATGGAAC 628
```

RESULT 10
AQ780940
LOCUS AQ780940 689 bp DNA linear GSS 02-AUG-1999
DEFINITION HS_3138_B1_B06_T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3138 Col=11 Row=D, genomic survey sequence.
ACCESSION AQ780940
VERSION AQ780940.1 GI:5683900
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 689)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
PUBMED 10449764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3138 row: D column: 11
Seq primer: T7
Class: BAC ends
High quality sequence stop: 689.
FEATURES Location/Qualifiers
source 1. .689
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=3138 Col=11 Row=D"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"
ORIGIN

Query Match 69.7%; Score 21.6; DB 28; Length 689;
Best Local Similarity 85.7%; Pred. No. 6.7e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy          2 ACAATTCAAACATTGTTCAAACAAGGAA 29
              | |||| ||| ||||||||| ||||||||
Db        407 ATAATTTAAATATTGTTCACACAAGGAA 434
```

RESULT 11
AZ202685/c
LOCUS AZ202685 734 bp DNA linear GSS 31-AUG-2000
DEFINITION SP_0059_B1_F01_T7A Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=59 Col=1 Row=L, genomic survey sequence.
ACCESSION AZ202685
VERSION AZ202685.1 GI:8397603
KEYWORDS GSS.
SOURCE Strongylocentrotus purpuratus
ORGANISM Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.

```
REFERENCE   1  (bases 1 to 734)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T.,
            Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H.
            and Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
   PUBMED   10920195
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 59  row: L  column: 1
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 734.
FEATURES             Location/Qualifiers
     source          1. .734
                     /organism="Strongylocentrotus purpuratus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7668"
                     /clone="Plate=59 Col=1 Row=L"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
ORIGIN

  Query Match             69.7%;  Score 21.6;  DB 28;  Length 734;
  Best Local Similarity   85.7%;  Pred. No. 6.7e+02;
  Matches   24;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAA 29
              ||||||||| | |||||||||| || |||
Db        333 ACAATTCAACCTTTGTTCAAAGAAAGAA 306


RESULT 12
AV961084/c
LOCUS       AV961084                283 bp    mRNA    linear   EST 14-MAR-2002
DEFINITION  AV961084 Nori Satoh unpublished cDNA library, cleavage stage embryo
            Ciona intestinalis cDNA clone cicl14p10 5', mRNA sequence.
ACCESSION   AV961084
VERSION     AV961084.1  GI:19449383
KEYWORDS    EST.
SOURCE      Ciona intestinalis
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 283)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .283
                     /organism="Ciona intestinalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:7719"
                     /clone="cicl14p10"
                     /tissue_type="whole animal"
                     /dev_stage="cleavage stage embryo"
                     /clone_lib="Nori Satoh unpublished cDNA library, cleavage
                     stage embryo".
ORIGIN

  Query Match             69.0%;  Score 21.4;  DB 9;  Length 283;
  Best Local Similarity   80.6%;  Pred. No. 8.3e+02;
  Matches   25;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGAACC 31
              | |||||| | |||||||  |||||| ||||
Db         33 TCCAATTCCACCATTGTTACAACAAGCAACC 3


RESULT 13
BI161768
LOCUS       BI161768                407 bp    mRNA    linear   EST 09-JUL-2001
DEFINITION  RE01178.3prime RE Drosophila melanogaster normalized Embryo pFlc-1
            Drosophila melanogaster cDNA clone RE01178 3, mRNA sequence.
ACCESSION   BI161768
VERSION     BI161768.1  GI:14627574
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
            Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
            George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
            Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
            Rubin,G.M.
  TITLE     BDGP/HHMI RE Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: RE01178.5prime
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003551: arm:3L [9396216,9682193]
            estimated-cyto:67B5-67C1: 04/11/2001
            Plate: RE.11  row: G  column: 6
            High quality sequence stop: 405.
FEATURES             Location/Qualifiers
     source          1. .407
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="RE01178"
                     /sex="male and female"
                     /dev_stage="0-24 hours mixed stage embryonic"
                     /lab_host="DH5-alpha TonA"
                     /clone_lib="RE Drosophila melanogaster normalized Embryo
                     pFlc-1"
                     /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library  was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
ORIGIN

  Query Match             69.0%;  Score 21.4;  DB 12;  Length 407;
  Best Local Similarity   80.6%;  Pred. No. 8.1e+02;
  Matches   25;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGAACC 31
              |||| || ||||||| |||||| ||| || |
Db         75 TACACTTTAAACATTTTTCAAAAAAGTAAAC 105


RESULT 14
```

```
BW195405/c
LOCUS       BW195405                486 bp    mRNA    linear   EST 05-NOV-2002
DEFINITION  BW195405 Nori Satoh unpublished cDNA library, cleaving embryo Ciona
            intestinalis cDNA clone cicl036c14 5', mRNA sequence.
ACCESSION   BW195405
VERSION     BW195405.1  GI:24609807
KEYWORDS    EST.
SOURCE      Ciona intestinalis
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.
  TITLE     Expressed genes in Ciona intestinalis (2002c)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="Ciona intestinalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:7719"
                     /clone="cicl036c14"
                     /tissue_type="whole body"
                     /dev_stage="cleaving embryo"
                     /clone_lib="Nori Satoh unpublished cDNA library, cleaving
                     embryo"
ORIGIN

  Query Match             69.0%;  Score 21.4;  DB 13;  Length 486;
  Best Local Similarity   80.6%;  Pred. No. 8.1e+02;
  Matches   25;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGAACC 31
              | |||||| | |||||||  |||||| ||||
Db         36 TCCAATTCCACCATTGTTACAACAAGCAACC 6


RESULT 15
BW209523/c
LOCUS       BW209523                486 bp    mRNA    linear   EST 05-NOV-2002
DEFINITION  BW209523 Nori Satoh unpublished cDNA library, egg Ciona
            intestinalis cDNA clone cieg061g07 5', mRNA sequence.
ACCESSION   BW209523
VERSION     BW209523.1  GI:24624137
KEYWORDS    EST.
SOURCE      Ciona intestinalis
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.
  TITLE     Expressed genes in Ciona intestinalis (2002c)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="Ciona intestinalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:7719"
                     /clone="cieg061g07"
                     /tissue_type="whole animal"
                     /dev_stage="egg"
                     /clone_lib="Nori Satoh unpublished cDNA library, egg"
ORIGIN

  Query Match             69.0%;  Score 21.4;  DB 13;  Length 486;
  Best Local Similarity   80.6%;  Pred. No. 8.1e+02;
  Matches   25;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGAACC 31
              | |||||| | |||||||  |||||| ||||
Db         36 TCCAATTCCACCATTGTTACAACAAGCAACC 6


Search completed: July 20, 2004, 01:42:49
Job time : 1424.41 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 00:23:29 ; Search time 172.279 Seconds
(without alignments)
877.864 Million cell updates/sec

Title: US-09-831-272-11
Perfect score: 31
Sequence: 1 tacaattcaaacattgttcaaacaaggaacc 31

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3190992 seqs, 2439311697 residues

Total number of hits satisfying chosen parameters: 6381984

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	21.6	69.7	91352	17	US-10-300-611-4	Sequence 4, Appli
	2	21	67.7	591	13	US-10-027-632-41058	Sequence 41058, A
	3	21	67.7	591	13	US-10-027-632-63246	Sequence 63246, A
	4	21	67.7	591	13	US-10-027-632-63857	Sequence 63857, A
	5	21	67.7	591	13	US-10-027-632-179516	Sequence 179516,
	6	21	67.7	591	13	US-10-027-632-310281	Sequence 310281,
	7	21	67.7	591	16	US-10-027-632-41058	Sequence 41058, A
	8	21	67.7	591	16	US-10-027-632-63246	Sequence 63246, A
	9	21	67.7	591	16	US-10-027-632-63857	Sequence 63857, A
	10	21	67.7	591	16	US-10-027-632-179516	Sequence 179516,
	11	21	67.7	591	16	US-10-027-632-310281	Sequence 310281,
	12	20.8	67.1	638	13	US-10-027-632-271776	Sequence 271776,
	13	20.8	67.1	638	16	US-10-027-632-271776	Sequence 271776,
	14	20.8	67.1	871	13	US-10-027-632-163062	Sequence 163062,
	15	20.8	67.1	871	13	US-10-027-632-163063	Sequence 163063,
	16	20.8	67.1	871	16	US-10-027-632-163062	Sequence 163062,
	17	20.8	67.1	871	16	US-10-027-632-163063	Sequence 163063,
c	18	20.4	65.8	1691139	15	US-10-067-514-1	Sequence 1, Appli
c	19	20.4	65.8	1691139	16	US-10-419-723-1	Sequence 1, Appli
	20	20.2	65.2	342	13	US-10-335-977-4117	Sequence 4117, Ap
	21	20.2	65.2	345	13	US-10-335-977-4116	Sequence 4116, Ap
	22	20.2	65.2	363	13	US-10-335-977-4118	Sequence 4118, Ap
	23	20.2	65.2	399	13	US-10-282-122A-41814	Sequence 41814, A
c	24	20.2	65.2	1253	13	US-10-424-599-135686	Sequence 135686,
c	25	20.2	65.2	84675	17	US-10-322-281-398	Sequence 398, App
	26	20.2	65.2	2940917	13	US-10-027-632-174763	Sequence 174763,
	27	20.2	65.2	2940917	16	US-10-027-632-174763	Sequence 174763,
c	28	20	64.5	406	10	US-09-918-995-3188	Sequence 3188, Ap
c	29	20	64.5	545	13	US-10-027-632-226281	Sequence 226281,
c	30	20	64.5	545	16	US-10-027-632-226281	Sequence 226281,
c	31	20	64.5	558	9	US-09-795-668-113	Sequence 113, App
c	32	20	64.5	558	9	US-09-795-686-113	Sequence 113, App
c	33	20	64.5	558	9	US-09-946-807-113	Sequence 113, App
c	34	20	64.5	612	13	US-10-027-632-271379	Sequence 271379,
c	35	20	64.5	612	13	US-10-027-632-271380	Sequence 271380,
c	36	20	64.5	612	16	US-10-027-632-271379	Sequence 271379,
c	37	20	64.5	612	16	US-10-027-632-271380	Sequence 271380,
c	38	20	64.5	747	13	US-10-027-632-271378	Sequence 271378,
c	39	20	64.5	747	16	US-10-027-632-271378	Sequence 271378,
	40	20	64.5	874	16	US-10-437-038-7	Sequence 7, Appli
	41	20	64.5	874	16	US-10-437-038-8	Sequence 8, Appli
	42	20	64.5	874	16	US-10-359-091-7	Sequence 7, Appli
	43	20	64.5	874	16	US-10-359-091-8	Sequence 8, Appli
c	44	20	64.5	2082	16	US-10-369-493-46177	Sequence 46177, A
	45	20	64.5	5092	9	US-09-834-975-1002	Sequence 1002, Ap

ALIGNMENTS

RESULT 1
US-10-300-611-4/c
; Sequence 4, Application US/10300611
; Publication No. US20040097451A1
; GENERAL INFORMATION:
; APPLICANT: Ming-Yi Chiang
; APPLICANT: Kenneth W. Dobie
; TITLE OF INVENTION: MODULATION OF NIDOGEN EXPRESSION
; FILE REFERENCE: PTS-0059
; CURRENT APPLICATION NUMBER: US/10/300,611
; CURRENT FILING DATE: 2002-11-19
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 4
; LENGTH: 91352
; TYPE: DNA
; ORGANISM: H. sapiens
US-10-300-611-4

Query Match 69.7%; Score 21.6; DB 17; Length 91352;
Best Local Similarity 85.7%; Pred. No. 5.5e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy          2 ACAATTCAAACATTGTTCAAACAAGGAA 29
              ||||||||  |||||||| ||||| ||
Db      86702 ACAATTCATTCATTGTTCACACAAGAAA 86675
```

RESULT 2
US-10-027-632-41058
; Sequence 41058, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632

```
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 41058
;   LENGTH: 591
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-41058

  Query Match             67.7%;  Score 21;  DB 13;  Length 591;
  Best Local Similarity   82.8%;  Pred. No. 3.3e+02;
  Matches   24;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAAC 30
              | |||| ||| |||| ||| |||||||||
Db        489 AAAATTGAAAAATTGATCAGACAAGGAAC 517


RESULT 3
US-10-027-632-63246
; Sequence 63246, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 63246
;   LENGTH: 591
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-63246

  Query Match             67.7%;  Score 21;  DB 13;  Length 591;
  Best Local Similarity   82.8%;  Pred. No. 3.3e+02;
  Matches   24;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAAC 30
              | |||| ||| |||| ||| |||||||||
Db        489 AAAATTGAAAAATTGATCAGACAAGGAAC 517


RESULT 4
US-10-027-632-63857
; Sequence 63857, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 63857
;   LENGTH: 591
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-63857

  Query Match             67.7%;  Score 21;  DB 13;  Length 591;
  Best Local Similarity   82.8%;  Pred. No. 3.3e+02;
  Matches   24;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAAC 30
              | |||| ||| |||| ||| |||||||||
Db        489 AAAATTGAAAAATTGATCAGACAAGGAAC 517


RESULT 5
US-10-027-632-179516
; Sequence 179516, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 179516
;   LENGTH: 591
;   TYPE: DNA
```

```
;   ORGANISM: Human
US-10-027-632-179516

  Query Match             67.7%;  Score 21;  DB 13;  Length 591;
  Best Local Similarity   82.8%;  Pred. No. 3.3e+02;
  Matches   24;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAAC 30
              | |||| ||| |||| ||| |||||||||
Db        489 AAAATTGAAAAATTGATCAGACAAGGAAC 517


RESULT 6
US-10-027-632-310281
; Sequence 310281, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 310281
;   LENGTH: 591
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-310281

  Query Match             67.7%;  Score 21;  DB 13;  Length 591;
  Best Local Similarity   82.8%;  Pred. No. 3.3e+02;
  Matches   24;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAAC 30
              | |||| ||| |||| ||| |||||||||
Db        489 AAAATTGAAAAATTGATCAGACAAGGAAC 517


RESULT 7
US-10-027-632-41058
; Sequence 41058, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 41058
;   LENGTH: 591
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-41058

  Query Match             67.7%;  Score 21;  DB 16;  Length 591;
  Best Local Similarity   82.8%;  Pred. No. 3.3e+02;
  Matches   24;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAAC 30
              | |||| ||| |||| ||| |||||||||
Db        489 AAAATTGAAAAATTGATCAGACAAGGAAC 517


RESULT 8
US-10-027-632-63246
; Sequence 63246, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 63246
;   LENGTH: 591
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-63246

  Query Match             67.7%;  Score 21;  DB 16;  Length 591;
  Best Local Similarity   82.8%;  Pred. No. 3.3e+02;
  Matches   24;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAAC 30
              | |||| ||| |||| ||| |||||||||
Db        489 AAAATTGAAAAATTGATCAGACAAGGAAC 517


RESULT 9
US-10-027-632-63857
; Sequence 63857, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
```

```
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 63857
;   LENGTH: 591
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-63857

  Query Match             67.7%;  Score 21;  DB 16;  Length 591;
  Best Local Similarity   82.8%;  Pred. No. 3.3e+02;
  Matches   24;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAAC 30
              | |||| ||| |||| ||| |||||||||
Db        489 AAAATTGAAAAATTGATCAGACAAGGAAC 517


RESULT 10
US-10-027-632-179516
; Sequence 179516, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 179516
;   LENGTH: 591
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-179516

  Query Match             67.7%;  Score 21;  DB 16;  Length 591;
  Best Local Similarity   82.8%;  Pred. No. 3.3e+02;
  Matches   24;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAAC 30
              | |||| ||| |||| ||| |||||||||
Db        489 AAAATTGAAAAATTGATCAGACAAGGAAC 517


RESULT 11
US-10-027-632-310281
; Sequence 310281, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 310281
;   LENGTH: 591
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-310281

  Query Match             67.7%;  Score 21;  DB 16;  Length 591;
  Best Local Similarity   82.8%;  Pred. No. 3.3e+02;
  Matches   24;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAAC 30
              | |||| ||| |||| ||| |||||||||
Db        489 AAAATTGAAAAATTGATCAGACAAGGAAC 517


RESULT 12
US-10-027-632-271776
; Sequence 271776, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 271776
```

```
;   LENGTH: 638
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-271776

  Query Match             67.1%;  Score 20.8;  DB 13;  Length 638;
  Best Local Similarity   91.7%;  Pred. No. 4e+02;
  Matches   22;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          6 TTCAAACATTGTTCAAACAAGGAA 29
              |||||||  ||||||||||||||||
Db         55 TTCAAACCGTGTTCAAACAAGGAA 78


RESULT 13
US-10-027-632-271776
; Sequence 271776, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 271776
;   LENGTH: 638
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-271776

  Query Match             67.1%;  Score 20.8;  DB 16;  Length 638;
  Best Local Similarity   91.7%;  Pred. No. 4e+02;
  Matches   22;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          6 TTCAAACATTGTTCAAACAAGGAA 29
              |||||||  ||||||||||||||||
Db         55 TTCAAACCGTGTTCAAACAAGGAA 78


RESULT 14
US-10-027-632-163062
; Sequence 163062, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 163062
;   LENGTH: 871
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-163062

  Query Match             67.1%;  Score 20.8;  DB 13;  Length 871;
  Best Local Similarity   91.7%;  Pred. No. 4.3e+02;
  Matches   22;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          6 TTCAAACATTGTTCAAACAAGGAA 29
              |||||||  ||||||||||||||||
Db         50 TTCAAACCGTGTTCAAACAAGGAA 73


RESULT 15
US-10-027-632-163063
; Sequence 163063, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 163063
;   LENGTH: 871
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-163063

  Query Match             67.1%;  Score 20.8;  DB 13;  Length 871;
  Best Local Similarity   91.7%;  Pred. No. 4.3e+02;
  Matches   22;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          6 TTCAAACATTGTTCAAACAAGGAA 29
              |||||||  ||||||||||||||||
Db         50 TTCAAACCGTGTTCAAACAAGGAA 73


Search completed: July 20, 2004, 01:48:41
Job time : 175.279 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 00:16:33 ; Search time 36.5902 Seconds
(without alignments)
470.167 Million cell updates/sec

Title: US-09-831-272-11
Perfect score: 31
Sequence: 1 tacaattcaaacattgttcaaacaaggaacc 31

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		20	64.5	3967	4	US-09-672-749-14	Sequence 14, Appl
c 2		19.6	63.2	750	4	US-09-328-352-1167	Sequence 1167, Ap
c 3		19.6	63.2	13868	4	US-08-956-171E-173	Sequence 173, App
c 4		19.6	63.2	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c 5		19.4	62.6	3396	3	US-08-974-549A-642	Sequence 642, App
c 6		19.4	62.6	3396	4	US-09-721-456-642	Sequence 642, App
c 7		19.4	62.6	58407	4	US-08-916-421B-2	Sequence 2, Appli
c 8		18.8	60.6	831	3	US-08-998-416-362	Sequence 362, App
9		18.8	60.6	942	4	US-08-956-171E-531	Sequence 531, App
c 10		18.8	60.6	3314	4	US-09-536-882A-5	Sequence 5, Appli
11		18.8	60.6	4594	4	US-08-956-171E-200	Sequence 200, App
12		18.8	60.6	10207	1	US-08-920-812-2	Sequence 2, Appli
13		18.8	60.6	10207	1	US-08-920-827-2	Sequence 2, Appli
14		18.8	60.6	10207	1	US-08-921-177-2	Sequence 2, Appli
15		18.8	60.6	10207	1	US-08-362-577C-2	Sequence 2, Appli
16		18.8	60.6	10207	2	US-08-920-828-2	Sequence 2, Appli
c 17		18.8	60.6	41708	4	US-09-470-512A-3	Sequence 3, Appli
18		18.8	60.6	1664976	4	US-08-916-421B-1	Sequence 1, Appli
19		18.6	60.0	9821	4	US-08-956-171E-470	Sequence 470, App
c 20		18.6	60.0	19932	2	US-08-477-451-25	Sequence 25, Appl
21		18.6	60.0	319608	4	US-09-539-333D-1	Sequence 1, Appli
22		18.6	60.0	319608	4	US-09-679-409-1	Sequence 1, Appli
23		18.4	59.4	125	1	US-07-998-003A-76	Sequence 76, Appl
24		18.4	59.4	125	1	US-08-453-274B-76	Sequence 76, Appl
25		18.4	59.4	125	1	US-08-453-695A-76	Sequence 76, Appl
26		18.4	59.4	125	1	US-08-268-161A-76	Sequence 76, Appl
27		18.4	59.4	125	2	US-08-453-702A-76	Sequence 76, Appl
28		18.4	59.4	125	3	US-09-099-639-76	Sequence 76, Appl
29		18.4	59.4	125	5	PCT-US93-12588-76	Sequence 76, Appl
30		18.4	59.4	125	5	PCT-US95-08071-76	Sequence 76, Appl
31		18.4	59.4	235	1	US-07-985-692-3	Sequence 3, Appli
32		18.4	59.4	235	1	US-08-155-331-3	Sequence 3, Appli
33		18.4	59.4	235	1	US-08-147-710-7	Sequence 7, Appli
34		18.4	59.4	235	1	US-08-321-658B-9	Sequence 9, Appli
35		18.4	59.4	235	1	US-08-384-489-6	Sequence 6, Appli
36		18.4	59.4	235	1	US-08-334-773A-3	Sequence 3, Appli
37		18.4	59.4	235	1	US-08-424-022-3	Sequence 3, Appli
38		18.4	59.4	235	1	US-08-458-090-7	Sequence 7, Appli
39		18.4	59.4	235	2	US-08-457-887-7	Sequence 7, Appli
40		18.4	59.4	235	2	US-08-424-017B-3	Sequence 3, Appli
41		18.4	59.4	235	4	US-09-904-621-7	Sequence 7, Appli
42		18.4	59.4	235	5	PCT-US93-11696-3	Sequence 3, Appli
43		18.4	59.4	348	1	US-08-468-674B-72	Sequence 72, Appl
44		18.4	59.4	348	1	US-08-780-571-72	Sequence 72, Appl
45		18.4	59.4	379	1	US-08-468-674B-74	Sequence 74, Appl

ALIGNMENTS

RESULT 1
US-09-672-749-14
; Sequence 14, Application US/09672749
; Patent No. 6475734
; GENERAL INFORMATION:
; APPLICANT: Liebergesell, Matthias
; APPLICANT: Fallis, Patricia L
; APPLICANT: Dong, Jian G
; APPLICANT: Li, Chun Ping
; APPLICANT: Nichols, Scott E
; TITLE OF INVENTION: POLYHYDROXYALKANOATE SYNTHASE GENES
; FILE REFERENCE: 5718-120
; CURRENT APPLICATION NUMBER: US/09/672,749
; CURRENT FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/156,770
; PRIOR FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 3967
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-672-749-14

Query Match 64.5%; Score 20; DB 4; Length 3967;
Best Local Similarity 82.1%; Pred. No. 46;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy          1 TACAATTCAAACATTGTTCAAACAAGGA 28
              ||| ||||||  || ||||||||||| |
Db       3895 TACGATTCAAGAATAGTTCAAACAAGAA 3922
```

RESULT 2
US-09-328-352-1167/c
; Sequence 1167, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1167
; LENGTH: 750
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1167

```
  Query Match             63.2%;  Score 19.6;  DB 4;  Length 750;
  Best Local Similarity   84.6%;  Pred. No. 57;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          5 ATTCAAACATTGTTCAAACAAGGAAC 30
              | | |||| |||||||||||||| |||
Db        321 ACTTAAACCTTGTTCAAACAAGCAAC 296


RESULT 3
US-08-956-171E-173/c
; Sequence 173, Application US/08956171E
; Patent No. 6593114
;   GENERAL INFORMATION:
;        APPLICANT: Charles Kunsch
;                   Gil H. Choi
;                   Patrick S. Dillon
;                   Craig A. Rosen
;                   Steven C. Barash
;                   Michael R. Fannon
;        TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
;        NUMBER OF SEQUENCES: 5256
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: Maryland
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;             COMPUTER: HP Vectra 486/33
;             OPERATING SYSTEM: MSDOS version 6.2
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/956,171E
;             FILING DATE: 20-Oct-1997
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/009,861
;             FILING DATE: January 5, 1996
;             APPLICATION NUMBER: 08/781,986
;             FILING DATE: January 3, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Mark J. Hyman
;             REGISTRATION NUMBER: 46,789
;             REFERENCE/DOCKET NUMBER: PB248P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (240) 314-1224
;             TELEFAX: (301) 309-8439
;   INFORMATION FOR SEQ ID NO: 173:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 13868 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 173:
US-08-956-171E-173

  Query Match             63.2%;  Score 19.6;  DB 4;  Length 13868;
  Best Local Similarity   84.6%;  Pred. No. 73;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          3 CAATTCAAACATTGTTCAAACAAGGA 28
              |||||| |||||||||| ||| |||
Db       5963 CAATTCCAACATTGTTCTTACATGGA 5938


RESULT 4
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococc
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
```

```
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (622708)..(622708)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657081)..(657081)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657203)..(657203)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (674435)..(674435)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (682442)..(682442)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (713652)..(713652)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (741684)..(741684)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779455)..(779455)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779676)..(779676)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (855539)..(855539)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (871619)..(871619)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1084830)..(1084830)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1096846)..(1096846)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1119881)..(1119881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1130881)..(1130881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1310988)..(1310988)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1313224)..(1313224)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349473)..(1349473)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349491)..(1349491)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1470091)..(1470091)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1569020)..(1569020)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1602912)..(1602912)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1603734)..(1603734)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1637998)..(1637998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1664854)..(1664855)
;   OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

  Query Match             63.2%;  Score 19.6;  DB 4;  Length 1664976;
  Best Local Similarity   84.6%;  Pred. No. 92;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAG 26
              ||||||||||| |||||  ||| ||
Db     187177 TACAATTCAAACTTTGTTATAACGAG 187152


RESULT 5
US-08-974-549A-642/c
; Sequence 642, Application US/08974549A
; Patent No. 6166178
;  GENERAL INFORMATION:
;    APPLICANT:  Cech, Thomas R.
;    APPLICANT:  Lingner, Joachim
;    APPLICANT:  Nakamura, Toru
;    APPLICANT:  Chapman, Karen B.
;    APPLICANT:  Morin, Gregg B.
;    APPLICANT:  Harley, Calvin B.
;    APPLICANT:  Andrews, William H.
;    TITLE OF INVENTION:  Human Telomerase Catalytic Subunit
;    NUMBER OF SEQUENCES:  727
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/974,549A
;      FILING DATE:  19-NOV-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/724,643
;      FILING DATE:  01-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/844,419
```

```
;       FILING DATE:  18-APR-1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/846,017
;       FILING DATE:  25-APR-1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/851,843
;       FILING DATE:  06-MAY-1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/854,050
;       FILING DATE:  09-MAY-1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/911,312
;       FILING DATE:  14-AUG-1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/912,951
;       FILING DATE:  14-AUG-1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/915,503
;       FILING DATE:  14-AUG-1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  WO PCT/US97/17618
;       FILING DATE:  01-OCT-1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  WO PCT/US97/17885
;       FILING DATE:  01-OCT-1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Apple, Randolph Ted
;       REGISTRATION NUMBER:  36,429
;       REFERENCE/DOCKET NUMBER:  015389-002610US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 576-0200
;       TELEFAX:  (415) 576-0300
;   INFORMATION FOR SEQ ID NO:  642:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3396 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
;     FEATURE:
;       NAME/KEY:  -
;       LOCATION:  1..3396
;       OTHER INFORMATION:  /note= ""generic" hTRT protein encoding
;       OTHER INFORMATION:  sequence"
US-08-974-549A-642

  Query Match             62.6%;  Score 19.4;  DB 3;  Length 3396;
  Best Local Similarity   79.3%;  Pred. No. 77;
  Matches   23;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAAC 30
              ||||  |||||||||  ||| ||||| ||
Db        171 ACAAACCAAACATTGGGCAACCAAGGCAC 143


RESULT 6
US-09-721-456-642/c
; Sequence 642, Application US/09721456
; Patent No. 6617110
;   GENERAL INFORMATION:
;        APPLICANT: Cech, Thomas R.
;                   Lingner, Joachim
;                   Nakamura, Toru
;                   Chapman, Karen B.
;                   Morin, Gregg B.
;                   Harley, Calvin B.
;                   Andrews, William H.
;        TITLE OF INVENTION: Human Telomerase Catalytic Subunit
;        NUMBER OF SEQUENCES: 727
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Townsend and Townsend and Crew LLP
;             STREET: Two Embarcadero Center, Eighth Floor
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94111-3834
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/721,456
;             FILING DATE: 22-No. 6617110-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/974,549A
;             FILING DATE: 19-NOV-1997
;             APPLICATION NUMBER: US 08/724,643
;             FILING DATE: 01-OCT-1996
;             APPLICATION NUMBER: US 08/844,419
;             FILING DATE: 18-APR-1997
;             APPLICATION NUMBER: US 08/846,017
;             FILING DATE: 25-APR-1997
;             APPLICATION NUMBER: US 08/851,843
;             FILING DATE: 06-MAY-1997
;             APPLICATION NUMBER: US 08/854,050
;             FILING DATE: 09-MAY-1997
;             APPLICATION NUMBER: US 08/911,312
;             FILING DATE: 14-AUG-1997
;             APPLICATION NUMBER: US 08/912,951
;             FILING DATE: 14-AUG-1997
;             APPLICATION NUMBER: US 08/915,503
;             FILING DATE: 14-AUG-1997
;             APPLICATION NUMBER: WO PCT/US97/17618
;             FILING DATE: 01-OCT-1997
;             APPLICATION NUMBER: WO PCT/US97/17885
;             FILING DATE: 01-OCT-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Apple, Randolph Ted
;             REGISTRATION NUMBER: 36,429
;             REFERENCE/DOCKET NUMBER: 015389-002610US
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 576-0200
;             TELEFAX: (415) 576-0300
;   INFORMATION FOR SEQ ID NO: 642:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 3396 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA
;        FEATURE:
;             NAME/KEY:  -
;             LOCATION:  1..3396
;             OTHER INFORMATION:  /note= ""generic" hTRT protein encoding
;             sequence"
;        SEQUENCE DESCRIPTION: SEQ ID NO: 642:
US-09-721-456-642

  Query Match             62.6%;  Score 19.4;  DB 4;  Length 3396;
  Best Local Similarity   79.3%;  Pred. No. 77;
  Matches   23;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAAC 30
              ||||  |||||||||  ||| ||||| ||
Db        171 ACAAACCAAACATTGGGCAACCAAGGCAC 143


RESULT 7
US-08-916-421B-2/c
; Sequence 2, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococc
```

```
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 58407
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6485)..(6485)
;   OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-2

  Query Match             62.6%;  Score 19.4;  DB 4;  Length 58407;
  Best Local Similarity   79.3%;  Pred. No. 97;
  Matches   23;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGAA 29
              || |||| | ||||| |||||| ||| ||
Db      13255 TATAATTTAGACATTCTTCAAAAAAGTAA 13227


RESULT 8
US-08-998-416-362/c
; Sequence 362, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  362:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  831 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1279RP
US-08-998-416-362

  Query Match             60.6%;  Score 18.8;  DB 3;  Length 831;
  Best Local Similarity   74.2%;  Pred. No. 1.2e+02;
  Matches   23;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGAACC 31
              || ||  ||||  ||| |||| ||||||| |
Db        825 TAAAAANCAAAAGTTGCTCAATCAAGGAAAC 795


RESULT 9
US-08-956-171E-531
; Sequence 531, Application US/08956171E
; Patent No. 6593114
;  GENERAL INFORMATION:
;       APPLICANT: Charles Kunsch
;                  Gil H. Choi
;                  Patrick S. Dillon
;                  Craig A. Rosen
;                  Steven C. Barash
;                  Michael R. Fannon
;       TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
;       NUMBER OF SEQUENCES: 5256
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville
;            STATE: Maryland
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;            COMPUTER: HP Vectra 486/33
;            OPERATING SYSTEM: MSDOS version 6.2
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/956,171E
;            FILING DATE: 20-Oct-1997
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/009,861
;            FILING DATE: January 5, 1996
;            APPLICATION NUMBER: 08/781,986
;            FILING DATE: January 3, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Mark J. Hyman
;            REGISTRATION NUMBER: 46,789
;            REFERENCE/DOCKET NUMBER: PB248P1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (240) 314-1224
;            TELEFAX: (301) 309-8439
;  INFORMATION FOR SEQ ID NO: 531:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 942 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 531:
US-08-956-171E-531

  Query Match             60.6%;  Score 18.8;  DB 4;  Length 942;
  Best Local Similarity   76.7%;  Pred. No. 1.2e+02;
  Matches   23;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGAAC 30
              | ||||  || |  |||| ||||||||||
Db        535 TTCAATGGAACCTATGTTAAAACAAGGAAC 564
```

```
RESULT 10
US-09-536-882A-5/c
; Sequence 5, Application US/09536882A
; Patent No. 6489151
; GENERAL INFORMATION:
;  APPLICANT: The Research Foundation of State University of New
;  TITLE OF INVENTION: A BIOLOGICALLY ACTIVE ALTERNATIVE FORM OF THE IKKa IkB
;  TITLE OF INVENTION:  Kinase
;  FILE REFERENCE: Seq. List 1-20 178-282
;  CURRENT APPLICATION NUMBER: US/09/536,882A
;  CURRENT FILING DATE:  2000-03-27
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 3314
;   TYPE: DNA
;   ORGANISM: Mouse
US-09-536-882A-5

  Query Match             60.6%;  Score 18.8;  DB 4;  Length 3314;
  Best Local Similarity   76.7%;  Pred. No. 1.3e+02;
  Matches   23;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAACC 31
              | |  | ||| |||||| ||||||| ||||
Db       2924 AGACATAAAAGATTGTTTAAACAAGTAACC 2895


RESULT 11
US-08-956-171E-200
; Sequence 200, Application US/08956171E
; Patent No. 6593114
;  GENERAL INFORMATION:
;       APPLICANT: Charles Kunsch
;                  Gil H. Choi
;                  Patrick S. Dillon
;                  Craig A. Rosen
;                  Steven C. Barash
;                  Michael R. Fannon
;       TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
;       NUMBER OF SEQUENCES: 5256
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville
;            STATE: Maryland
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;            COMPUTER: HP Vectra 486/33
;            OPERATING SYSTEM: MSDOS version 6.2
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/956,171E
;            FILING DATE: 20-Oct-1997
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/009,861
;            FILING DATE: January 5, 1996
;            APPLICATION NUMBER: 08/781,986
;            FILING DATE: January 3, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Mark J. Hyman
;            REGISTRATION NUMBER: 46,789
;            REFERENCE/DOCKET NUMBER: PB248P1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (240) 314-1224
;            TELEFAX: (301) 309-8439
;  INFORMATION FOR SEQ ID NO: 200:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 4594 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 200:
US-08-956-171E-200

  Query Match             60.6%;  Score 18.8;  DB 4;  Length 4594;
  Best Local Similarity   90.9%;  Pred. No. 1.3e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 AATTCAAACATTGTTCAAACAA 25
              |||||| |||||||||| |||
Db       1150 AATTCAATCATTGTTCAACCAA 1171


RESULT 12
US-08-920-812-2
; Sequence 2, Application US/08920812
; Patent No. 5763188
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,812
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10207 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Staphylococcus aureus
;      STRAIN:  Clinical Isolate SA-24
US-08-920-812-2

  Query Match             60.6%;  Score 18.8;  DB 1;  Length 10207;
  Best Local Similarity   90.9%;  Pred. No. 1.4e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 AATTCAAACATTGTTCAAACAA 25
              |||||| |||||||||| |||
Db       3737 AATTCAATCATTGTTCAACCAA 3758
```

```
RESULT 13
US-08-920-827-2
; Sequence 2, Application US/08920827
; Patent No. 5770375
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,827
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10207 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Staphylococcus aureus
;      STRAIN:  Clinical Isolate SA-24
US-08-920-827-2

  Query Match             60.6%;  Score 18.8;  DB 1;  Length 10207;
  Best Local Similarity   90.9%;  Pred. No. 1.4e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 AATTCAAACATTGTTCAAACAA 25
              ||||||| |||||||||| |||
Db       3737 AATTCAATCATTGTTCAACCAA 3758


RESULT 14
US-08-921-177-2
; Sequence 2, Application US/08921177
; Patent No. 5798211
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/921,177
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10207 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Staphylococcus aureus
;      STRAIN:  Clinical Isolate SA-24
US-08-921-177-2

  Query Match             60.6%;  Score 18.8;  DB 1;  Length 10207;
  Best Local Similarity   90.9%;  Pred. No. 1.4e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 AATTCAAACATTGTTCAAACAA 25
              ||||||| |||||||||| |||
Db       3737 AATTCAATCATTGTTCAACCAA 3758


RESULT 15
US-08-362-577C-2
; Sequence 2, Application US/08362577C
; Patent No. 5807673
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/362,577C
;      FILING DATE:  27-MAR-1995
```

; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: Clinical Isolate SA-24
US-08-362-577C-2

Query Match 60.6%; Score 18.8; DB 1; Length 10207;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          4 AATTCAAACATTGTTCAAACAA 25
              ||||||| |||||||||| |||
Db       3737 AATTCAATCATTGTTCAACCAA 3758
```

Search completed: July 20, 2004, 00:55:59
Job time : 38.5902 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 00:15:28 ; Search time 170.246 Seconds
(without alignments)
773.552 Million cell updates/sec

Title: US-09-831-272-11
Perfect score: 31
Sequence: 1 tacaattcaaacattgttcaaacaaggaacc 31

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	31	100.0	31	3	AAA27974	Aaa27974 Box D str
	2	21.4	69.0	2000	7	ACC61568	Acc61568 Gene sequ
c	3	21.4	69.0	4341	4	ABL20396	Abl20396 Drosophil
c	4	21.4	69.0	12010	4	ABL06138	Abl06138 Drosophil
	5	21.2	68.4	753	3	AAC33728	Aac33728 Arabidops
c	6	21	67.7	9507	6	ABN80242	Abn80242 Human che
c	7	20.4	65.8	541	5	ABV61041	Abv61041 Human pro
c	8	20.4	65.8	110000	6	ABX08336_03	Continuation (4 of
	9	20.2	65.2	342	2	AAT67714	Aat67714 H. pylori
	10	20.2	65.2	363	2	AAT67949	Aat67949 H. pylori
	11	20.2	65.2	399	7	ACA53944	Aca53944 Prokaryot
c	12	20	64.5	184	3	AAC12933	Aac12933 Human sec
c	13	20	64.5	406	8	ACH15976	Ach15976 Human adu
c	14	20	64.5	558	4	AAK95314	Aak95314 Human NRG
c	15	20	64.5	558	4	AAK96807	Aak96807 Human NRG
c	16	20	64.5	558	6	ABT00084	Abt00084 Human neu
c	17	20	64.5	558	6	ABT01577	Abt01577 Human neu
	18	20	64.5	658	1	AAN60062	Aan60062 Sequence
	19	20	64.5	814	4	AAA54544	Aaa54544 Human pre
	20	20	64.5	844	1	AAN50285	Aan50285 DNA encod
	21	20	64.5	848	1	AAN81387	Aan81387 Sequence
	22	20	64.5	873	2	AAQ90324	Aaq90324 MF-alpha-
	23	20	64.5	874	1	AAN80450	Aan80450 Sequence
	24	20	64.5	874	1	AAN80451	Aan80451 Sequence
	25	20	64.5	874	2	AAQ90325	Aaq90325 MF-alpha-
	26	20	64.5	874	4	AAA54543	Aaa54543 MF1::Huma
	27	20	64.5	1102	1	AAN40048	Aan40048 Sequence
c	28	20	64.5	2000	7	ACC60957	Acc60957 Gene sequ
	29	20	64.5	3967	4	AAD03830	Aad03830 Saccharom
	30	20	64.5	3967	4	AAD07041	Aad07041 Saccharom
c	31	20	64.5	4370	9	ADC51947	Adc51947 Human pos
c	32	20	64.5	4941	9	ADC51946	Adc51946 Human pos
	33	20	64.5	5092	4	AAS61001	Aas61001 Human can
c	34	20	64.5	5575	6	AAS61207	Aas61207 Human gen
	35	20	64.5	6241	4	AAL04604	Aal04604 Human rep
	36	20	64.5	6241	4	ABL97527	Abl97527 Human tes
c	37	20	64.5	7525	4	ABL17686	Abl17686 Drosophil
c	38	20	64.5	110000	4	AAK95240_06	Continuation (7 of
c	39	20	64.5	110000	4	AAK96733_06	Continuation (7 of
c	40	20	64.5	110000	6	ABT00010_06	Continuation (7 of
c	41	20	64.5	110000	6	ABT01503_06	Continuation (7 of
	42	19.8	63.9	12791	4	ABL11557	Abl11557 Drosophil
	43	19.8	63.9	17282	4	ABL11556	Abl11556 Drosophil
c	44	19.6	63.2	365	9	ADE07038	Ade07038 Novel cod
	45	19.6	63.2	706	2	AAV81170	Aav81170 DNA seque

ALIGNMENTS

RESULT 1
AAA27974
ID AAA27974 standard; DNA; 31 BP.
XX
AC AAA27974;
XX
DT 15-AUG-2000 (first entry)
XX
DE Box D strong elicitor-responsive cis-element nucleotide sequence.
XX
KW Box D; elicitor-responsive cis-element; parsley; PR1 promoter; ss;
KW chimeric promoter; pathogen infection; transgenic plant; resistance;
KW herbicide; local response; genetic engineering; disease resistant crop.
XX
OS Petroselinum crispum.
XX
PN WO200029592-A2.
XX
PD 25-MAY-2000.
XX
PF 12-NOV-1999; 99WO-EP008710.
XX
PR 12-NOV-1998; 98EP-00121160.
PR 27-AUG-1999; 99EP-00116981.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG.
XX
PI Kirsch C, Logemann E, Hahlbrock K, Rushton P, Somssich I;
XX
DR WPI; 2000-387804/33.
XX
PT Chimeric promoters mediating gene expression in plants upon pathogen
PT infection, useful for transgenic plant production comprises at least one
PT cis-acting element to direct elicitor-specific expression.
XX
PS Claim 1; Page 30; 73pp; English.
XX
CC This sequence represents Box D, a strong elicitor-responsive cis-element
CC from the Parsley PR1 promoter. The present invention relates to chimeric
CC promoters capable of mediating local gene expression in plants upon
CC pathogen infection. The chimeric promoters comprise at least one cis-
CC element (see AAA27964-A27979) capable of directing elicitor-specific
CC expression, and a minimal promoter. The chimeric promoters are useful for
CC producing a transgenic plant which has attained resistance or improved
CC resistance against a pathogen. The cis-acting element, chimeric promoter,
CC recombinant gene encoding the chimeric promoter, vector comprising the

```
CC   chimeric promoter and a compound capable of activating the chimeric
CC   promoter are useful for producing pathogen resistant plants, and for
CC   identifying and/or producing compounds capable of conferring induced
CC   resistance to a pathogen in a plant. A compound which specifically
CC   activates or inhibits genes activated in a plant when attacked by a
CC   pathogen is also useful as a plant protective agent or a herbicide. The
CC   chimeric promoter provides rapid and local response to pathogen attack
CC   but shows negligible activity in uninfected parts of the plants and
CC   therefore is most suitable for the engineering of disease resistant crops
XX
SQ   Sequence 31 BP; 14 A; 7 C; 3 G; 7 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 31;  DB 3;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 0.021;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGAACC 31
              |||||||||||||||||||||||||||||||
Db          1 TACAATTCAAACATTGTTCAAACAAGGAACC 31


RESULT 2
ACC61568
ID   ACC61568 standard; DNA; 2000 BP.
XX
AC   ACC61568;
XX
DT   20-JUN-2003  (first entry)
XX
DE   Gene sequence #SEQ ID 1918.
XX
KW   Multiprotein complex; eukaryote; drug target; diagnosis; gene; ds.
XX
OS   Saccharomyces cerevisiae.
XX
PN   EP1258494-A1.
XX
PD   20-NOV-2002.
XX
PF   20-DEC-2001; 2001EP-00130253.
XX
PR   15-MAY-2001; 2001EP-00111774.
XX
PA   (CELL-) CELLZOME AG.
XX
PI   Bauer A,  Gavin A,  Grandi P,  Krause R,  Kruse UD,  Kuester BD;
PI   Marzioch M,  Schultz JD,  Superti-Furga GD;
XX
DR   WPI; 2003-250078/25.
DR   P-PSDB; ABR53526.
XX
PT   New isolated protein complexes useful for diagnosing a disease or
PT   disorder, or as a target for an active agent of a pharmaceutical,
PT   preferably a drug target in the treatment or prevention of disease or
PT   disorder.
XX
PS   Disclosure; SEQ ID NO 1918; 17pp + Sequence Listing; English.
XX
CC   The invention relates to multiprotein complexes from eukaryotes. Proteins
CC   of the invention and DNA sequences encoding them are given in records
CC   ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are
CC   obtainable by using a protein as a bait and isolating the set of proteins
CC   which is attached thereto from cells. Such protein complexes may comprise
CC   up to 30 distinct proteins. Protein complexes of the invention are useful
CC   for diagnosing a disease or disorder, or as a target for an active agent
CC   of a pharmaceutical, preferably a drug target in the treatment or
CC   prevention of a disease or disorder. Note: The sequence data for this
CC   patent is not represented in the printed specification, but is based on
CC   sequence information supplied by the European Patent Office. The complete
CC   document is available on CD-ROM
XX
SQ   Sequence 2000 BP; 768 A; 367 C; 320 G; 545 T; 0 U; 0 Other;

  Query Match             69.0%;  Score 21.4;  DB 7;  Length 2000;
  Best Local Similarity   80.6%;  Pred. No. 1.1e+02;
  Matches   25;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGAACC 31
              ||||| ||||| || ||||||| ||| || |
Db       1666 TACAAATCAAAAATGGTTCAAAAAAGAAAAC 1696


RESULT 3
ABL20396/c
ID   ABL20396 standard; DNA; 4341 BP.
XX
AC   ABL20396;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 12661.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 12661; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4341 BP; 1299 A; 863 C; 857 G; 1322 T; 0 U; 0 Other;

  Query Match             69.0%;  Score 21.4;  DB 4;  Length 4341;
  Best Local Similarity   80.6%;  Pred. No. 1.1e+02;
  Matches   25;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAAGGAACC 31
              |||| || ||||||| |||||| ||| || |
Db       4019 TACACTTTAAACATTTTTCAAAAAAGTAAAC 3989


RESULT 4
ABL06138/c
ID   ABL06138 standard; cDNA; 12010 BP.
XX
AC   ABL06138;
XX
```

DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 12896.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB62035.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 12896; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 12010 BP; 3743 A; 2254 C; 2323 G; 3690 T; 0 U; 0 Other;

Query Match 69.0%; Score 21.4; DB 4; Length 12010;
Best Local Similarity 80.6%; Pred. No. 1.2e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 TACAATTCAAACATTGTTCAAACAAGGAACC 31
|||| || ||||||| |||||| ||| || |
Db 11688 TACACTTTAAACATTTTTCAAAAAAGTAAAC 11658

RESULT 5
AAC33728
ID AAC33728 standard; DNA; 753 BP.
XX
AC AAC33728;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 4105.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.

PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 68.4%; Score 21.2; DB 3; Length 753;
Best Local Similarity 88.5%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          5 ATTCAAACATTGTTCAAACAAGGAAC 30
              ||||| | ||||||||||||||| ||
Db         41 ATTCAGAGATTGTTCAAACAAGGCAC 66
```

RESULT 6
ABN80242/c
ID ABN80242 standard; DNA; 9507 BP.
XX
AC ABN80242;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human chemically modified disease associated gene SEQ ID NO 259.
XX
KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
KW antidiabetic; cytostatic; anticonvulsant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200927-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007536.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX

PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130908/17.
XX
PT Novel nucleic acid useful for diagnosis and therapy of diseases
PT associated with development genes such as diabetes, comprises a sequence
PT of a segment of chemically pretreated DNA of genes associated with
PT development.
XX
PS Claim 1; SEQ ID NO 259; 27pp; English.
XX
CC The invention relates to a nucleic acid (I) comprising a sequence at
CC least 18 bases in length of a segment of chemically pretreated DNA (II)
CC of genes associated with development selected from 87 genes listed in the
CC specification such as ACCPN, ADFN, or AFD1 and comprising one of 350
CC sequences (ABN79984-ABN80333) or their complements. The invention is
CC useful for the diagnosis or therapy of diseases associated with
CC development genes, in particular disease related to homeobox containing
CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
CC associated with congenital heart disease, epilepsy, diseases related to
CC histone deacetylation, Currarino syndrome, diseases related with the
CC development of the brain and limb girdle muscular dystrophy and dwarfism.
CC Oligomers specific to each of the genes are useful for detecting the
CC methylation state of all CpG dinucleotides within the 350 sequences or
CC (II) and their complementary sequences, as primer oligonucleotides for
CC the amplification of the 350 sequences, (II) and/or their complements and
CC as oligomer probes for detecting the cytosine methylation state and/or
CC single nucleotide polymorphisms (SNPs). Note: The sequence data for this
CC patent did not form part of the printed specification but is based on
CC sequence information supplied to Derwent by the European Patent Office
XX
SQ Sequence 9507 BP; 2235 A; 326 C; 2497 G; 4449 T; 0 U; 0 Other;

Query Match 67.7%; Score 21; DB 6; Length 9507;
Best Local Similarity 82.8%; Pred. No. 1.7e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACAATTCAAACATTGTTCAAACAAGGAAC 30
||||||||||| | ||||||||| | |
Db 1899 ACAATTCAAACAATATTCAAACAATAACC 1871

RESULT 7
ABV61041/c
ID ABV61041 standard; cDNA; 541 BP.
XX
AC ABV61041;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 61032.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 11599; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ Sequence 541 BP; 205 A; 98 C; 100 G; 138 T; 0 U; 0 Other;

Query Match 65.8%; Score 20.4; DB 5; Length 541;
Best Local Similarity 80.0%; Pred. No. 2.5e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 ACAATTCAAACATTGTTCAAACAAGGAACC 31
|| | |||||| ||||||||| ||| |||
Db 77 ACTACTCAAACCTTGTTCAAATAAGACACC 48

RESULT 8
ABX08336_03/c
Continuation (4 of 17) of ABX08336 from base 300001 (Human phosphodiesterase 4D (PDE4D) s
WP Sequence split into 17 fragments LOCUS ABX08336 Accession Abx08336

WP	Fragment Name	Begin	End
WP	ABX08336_00	1	110000
WP	ABX08336_01	100001	210000
WP	ABX08336_02	200001	310000
WP	ABX08336_03	300001	410000
WP	ABX08336_04	400001	510000
WP	ABX08336_05	500001	610000
WP	ABX08336_06	600001	710000
WP	ABX08336_07	700001	810000
WP	ABX08336_08	800001	910000
WP	ABX08336_09	900001	1010000
WP	ABX08336_10	1000001	1110000
WP	ABX08336_11	1100001	1210000
WP	ABX08336_12	1200001	1310000
WP	ABX08336_13	1300001	1410000
WP	ABX08336_14	1400001	1510000
WP	ABX08336_15	1500001	1610000
WP	ABX08336_16	1600001	1691080

Query Match 65.8%; Score 20.4; DB 6; Length 110000;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 ACAATTCAAACATTGTTCAAACAAGGAACC 31
|| | |||||| ||||||||| ||| |||
Db 53790 ACTACTCAAACCTTGTTCAAATAAGACACC 53761

RESULT 9
AAT67714
ID AAT67714 standard; DNA; 342 BP.
XX
AC AAT67714;
XX

```
DT   17-JUL-1997  (first entry)
XX
DE   H. pylori secreted or periplasmic protein ORF 80257.aa.
XX
KW   Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW   duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .342
FT                   /*tag=  a
FT                   /note= "no stop codon given"
XX
PN   WO9640893-A1.
XX
PD   19-DEC-1996.
XX
PF   06-JUN-1996;   96WO-US009122.
XX
PR   07-JUN-1995;   95US-00487032.
PR   01-APR-1996;   96US-00630405.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Smith D,  Berglindh OT,  Mellgaerd BL;
XX
DR   WPI; 1997-052306/05.
DR   P-PSDB; AAW20571.
XX
PT   Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
PT   useful for vaccines to treat or prevent H. pylori infection, and to
PT   detect Helicobacter.
XX
PS   Claim 23; Page 317; 1481pp; English.
XX
CC   This sequence encodes a H. pylori secreted or periplasmic protein. The
CC   protein may be used in a vaccine to prevent or treat H. pylori infection
CC   or to identify H. pylori polypeptide binding compounds, useful as
CC   potential H. pylori life cycle activators or inhibitors. The genomic
CC   sequence of H. pylori (ATCC 55679) was determined from overlapping
CC   contigs generated by mechanically shearing the bacterial DNA. The
CC   sequences were analysed for ORF of at least 180 nucleotides, and the
CC   predicted coding regions defined by computer evaluation. To identify
CC   likely H. pylori antigens for vaccine development, the amino acid
CC   sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts
XX
SQ   Sequence 342 BP; 106 A; 64 C; 44 G; 128 T; 0 U; 0 Other;

  Query Match             65.2%;  Score 20.2;  DB 2;  Length 342;
  Best Local Similarity   88.0%;  Pred. No. 2.9e+02;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAA 25
              |||||||||||||||  ||  |||||
Db         38 TACAATTCAAACATTCTTTCAACAA 62


RESULT 10
AAT67949
ID   AAT67949 standard; DNA; 363 BP.
XX
AC   AAT67949;
XX
DT   15-JUL-1997  (first entry)
XX
DE   H. pylori secreted or periplasmic protein ORF 05ae20220orf50.
XX
KW   Secretion; vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW   duodenal ulcer disease; chronic gastritis; diagnosis; envelope;
KW   periplasm; ds.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .363
FT                   /*tag=  a
FT                   /note= "no stop codon given"
XX
PN   WO9640893-A1.
XX
PD   19-DEC-1996.
XX
PF   06-JUN-1996;   96WO-US009122.
XX
PR   07-JUN-1995;   95US-00487032.
PR   01-APR-1996;   96US-00630405.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Smith D,  Berglindh OT,  Mellgaerd BL;
XX
DR   WPI; 1997-052306/05.
DR   P-PSDB; AAW20696.
XX
PT   Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
PT   useful for vaccines to treat or prevent H. pylori infection, and to
PT   detect Helicobacter.
XX
PS   Claim 23; Page 806; 1481pp; English.
XX
CC   The present sequence encodes a Helicobacter pylori protein which is
CC   likely to be secreted or periplasmic. The protein may be used in a
CC   vaccine to prevent or treat H. pylori infection or to identify H. pylori
CC   polypeptide binding compounds, useful as potential H. pylori life cycle
CC   activators or inhibitors. The genomic sequence of H. pylori (ATCC 55679)
CC   was determined from overlapping contigs generated by mechanically
CC   shearing the bacterial DNA. The sequences were analysed for ORF of at
CC   least 180 nucleotides, and the predicted coding regions defined by
CC   computer evaluation. To identify likely H. pylori antigens for vaccine
CC   development, the amino acid sequences predicted from various ORF were
CC   analysed for significant homology to other known or exported membrane
CC   proteins. Having identified and determined the sequences of interest,
CC   particular regions can be isolated from H. pylori by PCR amplification
CC   for recombinant polypeptide production, e.g. in E. coli hosts
XX
SQ   Sequence 363 BP; 116 A; 69 C; 44 G; 134 T; 0 U; 0 Other;

  Query Match             65.2%;  Score 20.2;  DB 2;  Length 363;
  Best Local Similarity   88.0%;  Pred. No. 2.9e+02;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 TACAATTCAAACATTGTTCAAACAA 25
              |||||||||||||||  ||  |||||
Db         59 TACAATTCAAACATTCTTTCAACAA 83


RESULT 11
ACA53944
ID   ACA53944 standard; DNA; 399 BP.
XX
AC   ACA53944;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Prokaryotic essential gene #35601.
XX
KW   Antisense; ds; prokaryotic essential gene; cell proliferation;
KW   drug design; gene.
XX
```

OS Yersinia pestis.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU50074.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 41814; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 399 BP; 102 A; 98 C; 95 G; 104 T; 0 U; 0 Other;

Query Match 65.2%; Score 20.2; DB 7; Length 399;
Best Local Similarity 88.0%; Pred. No. 2.9e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 7 TCAAACATTGTTCAAACAAGGAACC 31
||| ||||||||| |||| ||||||
Db 363 TCATACATTGTTCCAACAGGGAACC 387

RESULT 12
AAC12933/c
ID AAC12933 standard; cDNA; 184 BP.
XX
AC AAC12933;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 17008.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-00200610.
XX
PR 26-FEB-1999; 99US-0122487P.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS Claim 1; SEQ ID NO 17008; 71pp + Sequence Listing; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors
XX
SQ Sequence 184 BP; 53 A; 36 C; 44 G; 51 T; 0 U; 0 Other;

Query Match 64.5%; Score 20; DB 3; Length 184;
Best Local Similarity 82.1%; Pred. No. 3.3e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACAATTCAAACATTGTTCAAACAAGGAA 29
|||||||| ||||||||| || || ||
Db 68 ACAATTCACACATTGTTCCAAGGAGTAA 41

RESULT 13
ACH15976/c
ID ACH15976 standard; cDNA; 406 BP.
XX
AC ACH15976;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human adult heart cDNA #290.
XX
KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.

XX
PN US2003073623-A1.
XX
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.
XX
PR 30-JUL-2001; 2001US-00918995.
XX
PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
DR WPI; 2003-615964/58.
XX
PT New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
PS Claim 1; SEQ ID NO 3188; 44pp; English.
XX
CC The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversities, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ Sequence 406 BP; 148 A; 70 C; 88 G; 100 T; 0 U; 0 Other;

```
  Query Match             64.5%;  Score 20;  DB 8;  Length 406;
  Best Local Similarity   82.1%;  Pred. No. 3.5e+02;
  Matches   23;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAA 29
              ||  |||||||  |||||||| ||||||
Db        388 ACTGTTCAAACTGTGTTCAAATAAGGAA 361
```

RESULT 14
AAK95314/c
ID AAK95314 standard; DNA; 558 BP.
XX
AC AAK95314;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human NRG1AG1 exon single nucleotide polymorphism E588B.
XX
KW Human; neuregulin-1 associated gene 1; NRG1AG1; Schizophrenia gene;
KW gene therapy; single nucleotide polymorphism; SNP; ds.
XX
OS Homo sapiens.
XX
PN WO200164876-A2.
XX
PD 07-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US006376.
XX
PR 28-FEB-2000; 2000US-00515715.
XX
PA (DECO-) DECODE GENETICS EHF.
XX
PI Stefansson H, Steinthorsdottir V, Gulcher JR;
XX
DR WPI; 2001-550179/61.
XX
PT Neuregulin-1 associated gene 1 nucleic acids and fragments, useful for
PT preventing diagnosing and treating schizophrenia.
XX
PS Disclosure; Page 507; 750pp; English.
XX
CC This sequence represents a single nucleotide polymorphism (SNP) of the
CC human neuregulin-1 associated gene 1 (NRG1AG1) of the invention. The
CC NRG1AG1 gene is also referred to as the human Schizophrenia gene. The
CC invention also relates to fragments or variants of the gene and the
CC NRG1AG1 polypeptides they encode. The NRG1AG1 nucleic acids and
CC polypeptides may be used in the prevention, diagnosis and treatment of
CC diseases associated with inappropriate NRG1AG1 expression. For example,
CC they may be used to treat disorders associated with decreased expression
CC by rectifying mutations or deletions in a patient's genome that affect
CC the activity of NRG1AG1 by expressing inactive proteins or to supplement
CC the patients own production of NRG1AG1. Additionally, the gene may be
CC used to produce NRG1AG1 polypeptides, by inserting the nucleic acids into
CC a host cell and culturing the cell to express the protein. The gene may
CC also be used as DNA probes and primers in diagnostic assays to detect and
CC quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. The
CC NRG1AG1 polypeptides may also be used as antigens in the production of
CC antibodies against NRG1AG1 and in assays to identify modulators of
CC NRG1AG1 expression and activity. Anti-NRG1AG1 antibodies and antagonists
CC may also be used to down regulate expression and activity. Anti-NRG1AG1
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of NRG1AG1 polypeptides in samples. NRG1AG1 Is associated with
CC schizophrenia which may be prevented, diagnosed and/or treated by the
CC above methods
XX
SQ Sequence 558 BP; 185 A; 81 C; 104 G; 187 T; 0 U; 1 Other;

```
  Query Match             64.5%;  Score 20;  DB 4;  Length 558;
  Best Local Similarity   82.1%;  Pred. No. 3.5e+02;
  Matches   23;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACAATTCAAACATTGTTCAAACAAGGAA 29
              ||  |||||||  |||||||||||| |
Db        175 ACTGTTCAAACCATGTTCAAACAAGGCA 148
```

RESULT 15
AAK96807/c
ID AAK96807 standard; DNA; 558 BP.
XX
AC AAK96807;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human NRG1AG1 exon single nucleotide polymorphism E588B.
XX
KW Human; neuregulin 1 gene; schizophrenia; gene therapy; SNP;
KW single nucleotide polymorphism; ds.
XX
OS Homo sapiens.
XX
PN WO200164877-A2.
XX
PD 07-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US006377.
XX

PR 28-FEB-2000; 2000US-00515716.
XX
PA (DECO-) DECODE GENETICS EHF.
XX
PI Stefansson H, Steinthorsdottir V, Gulcher JR;
XX
DR WPI; 2001-514841/56.
XX
PT Neuregulin 1 nucleic acids and proteins useful for diagnosing preventing
PT and treating schizophrenia.
XX
PS Disclosure; Page 92; 756pp; English.
XX
CC This sequence represents a single nucleotide polymorphism (SNP) from the
CC human neuregulin 1 gene of the invention. The invention also relates to
CC fragments or variants of the neuregulin 1 gene. The gene and its proteins
CC may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate neuregulin 1 expression, such as
CC schizophrenia. For example they may be used to treat disorders associated
CC with decreased neuregulin 1 expression by rectifying mutations or
CC deletions in a patient's genome that affect the activity of neuregulin 1
CC by expressing inactive proteins or to supplement the patients own
CC production of polypeptides. Additionally, the gene may be used to produce
CC the neuregulin 1 protein, by inserting the nucleic acids into a host cell
CC and culturing the cell to express the protein. The gene and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids in
CC samples, and therefore which patients may be in need of restorative
CC therapy. The protein may also be used as antigens in the production of
CC antibodies against neuregulin 1 and in assays to identify modulators of
CC neuregulin 1 expression and activity. The antibodies and antagonists may
CC also be used to down regulate expression and activity. The antibodies may
CC also be used as diagnostic agents for detecting the presence of
CC neuregulin 1 in samples
XX
SQ Sequence 558 BP; 185 A; 81 C; 104 G; 187 T; 0 U; 1 Other;

Query Match 64.5%; Score 20; DB 4; Length 558;
Best Local Similarity 82.1%; Pred. No. 3.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy          2 ACAATTCAAACATTGTTCAAACAAGGAA 29
              ||  |||||||  |||||||||||||| |
Db        175 ACTGTTCAAACCATGTTCAAACAAGGCA 148
```

Search completed: July 20, 2004, 00:29:14
Job time : 174.246 secs

This page intentionally left blank

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 00:15:58 ; Search time 745.574 Seconds
(without alignments)
1744.014 Million cell updates/sec

Title: US-09-831-272-2
Perfect score: 30
Sequence: 1 cacacttaatttgaccgagtaacattcgcc 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		30	100.0	30	6	AX025602	AX025602 Sequence
2		30	100.0	930	8	PCPR115	X12572 Parsley PcP
3		30	100.0	2011	8	PCU48862	U48862 Petroselinu
4		28.4	94.7	1312	8	PCU48863	U48863 Petroselinu
c 5		20.4	68.0	91811	10	AL670670	AL670670 Mouse DNA
c 6		20.4	68.0	94638	9	AL513220	AL513220 Human DNA
c 7		20.4	68.0	112224	9	HS423B22	AL034379 Human DNA
c 8		20.4	68.0	153023	2	AL355880	AL355880 Homo sapi
9		19.8	66.0	336609	1	BX571867	BX571867 Photorhab
c 10		19.8	66.0	349980	6	AX770906	AX770906 Sequence
11		19.6	65.3	59008	2	AC100379	AC100379 Mus muscu
c 12		19.6	65.3	60978	2	AC100388	AC100388 Mus muscu
13		19.6	65.3	103930	9	AL358234	AL358234 Human DNA
14		19.6	65.3	110000	2	AC026388_0	AC026388 Mus muscu
15		19.6	65.3	139228	4	AC127467	AC127467 Atelerix
c 16		19.6	65.3	176184	9	AC018845	AC018845 Homo sapi
17		19.6	65.3	182063	2	BX510924	BX510924 Danio rer
18		19.6	65.3	185281	9	AC007338	AC007338 Homo sapi
c 19		19.6	65.3	211818	5	BX649401	BX649401 Zebrafish
20		19.6	65.3	228585	2	BX547928	BX547928 Danio rer
c 21		19.6	65.3	242443	2	AC112594	AC112594 Rattus no
22		19.4	64.7	61	6	AR033143	AR033143 Sequence
23		19.4	64.7	61	6	I41223	I41223 Sequence 30
24		19.4	64.7	61	6	I92455	I92455 Sequence 30
25		19.4	64.7	619	14	AY091551	AY091551 Avian inf
26		19.4	64.7	143411	2	AC102183	AC102183 Mus muscu
c 27		19.4	64.7	150347	2	AC108184	AC108184 Felis cat
c 28		19.4	64.7	164317	9	AL390029	AL390029 Human DNA
c 29		19.4	64.7	215314	10	AC140373	AC140373 Mus muscu
c 30		19.4	64.7	221530	2	AC138095	AC138095 Mus muscu
31		19.4	64.7	228450	2	AC112432	AC112432 Rattus no
c 32		19.4	64.7	260430	2	AC097758	AC097758 Rattus no
c 33		19.4	64.7	319087	2	AC135770	AC135770 Rattus no
c 34		19.4	64.7	349980	6	AX344571	AX344571 Sequence
35		19.2	64.0	51536	3	AC004429	AC004429 Drosophil
c 36		19.2	64.0	53337	9	AL603632	AL603632 Human DNA
c 37		19.2	64.0	141337	2	BX511248	BX511248 Danio rer
38		19.2	64.0	148418	2	AC018258	AC018258 Drosophil
39		19.2	64.0	192358	3	AC099028	AC099028 Drosophil
c 40		19.2	64.0	222887	2	AC015681	AC015681 Homo sapi
c 41		19.2	64.0	232829	2	AC099126	AC099126 Rattus no
42		19.2	64.0	242365	3	AE003798	AE003798 Drosophil
43		19.2	64.0	302399	1	AE017164	AE017164 Prochloro
44		19	63.3	1901	5	BC048876	BC048876 Danio rer
c 45		19	63.3	14968	9	HSDNAIL15	X91233 H.sapiens I

ALIGNMENTS

RESULT 1
AX025602
LOCUS AX025602 30 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 2 from Patent WO0029592.
ACCESSION AX025602
VERSION AX025602.1 GI:10187270
KEYWORDS .
SOURCE Petroselinum crispum (parsley)
ORGANISM Petroselinum crispum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Apiales; Apiaceae; Apioideae; apioid superclade; Apium clade; Petroselinum.
REFERENCE 1
AUTHORS Logemann,E., Somssich,I., Hahlbrock,K., Kirsch,C. and Rushton,P.

```
  TITLE     Chimeric promoters capable of mediating gene expression in plants
            upon pathogen infection and uses thereof
  JOURNAL   Patent: WO 0029592-A 2 25-MAY-2000;
            MAX PLANCK GESELLSCHAFT (DE) ; LOGEMANN ELKE (DE) ; SOMSSICH IMRE
            (DE) ; HAHLBROCK KLAUS (DE) ; KIRSCH CHRISTOPH (DE) ; RUSHTON PAUL
            (GB)
FEATURES             Location/Qualifiers
     source          1. .30
                     /organism="Petroselinum crispum"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:4043"
ORIGIN

  Query Match           100.0%;  Score 30;  DB 6;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 0.0021;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATTCGCC 30
              ||||||||||||||||||||||||||||||
Db          1 CACACTTAATTTGACCGAGTAACATTCGCC 30


RESULT 2
PCPR115
LOCUS       PCPR115                   930 bp    DNA     linear   PLN 30-MAY-1989
DEFINITION  Parsley PcPR1-1 gene for pathogenesis-related protein type A 5'
            UTR.
ACCESSION   X12572
VERSION     X12572.1  GI:20455
KEYWORDS    pathogenesis-related protein.
SOURCE      Petroselinum crispum (parsley)
  ORGANISM  Petroselinum crispum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Apiales; Apiaceae; Apioideae; apioid
            superclade; Apium clade; Petroselinum.
REFERENCE   1  (bases 1 to 930)
  AUTHORS   Somssich,I.E., Schmelzer,E., Kawalleck,P. and Hahlbrock,K.
  TITLE     Gene structure and in situ transcript localization of
            pathogenesis-related protein 1 in parsley
  JOURNAL   Mol. Gen. Genet. 213 (1), 93-98 (1988)
  MEDLINE   89127150
   PUBMED   3221838
COMMENT     see x15085 for PR1-1 gene 3' UTR; see x12573 and x12574 for PR1-3
            and PR1-2 cDNA seqs;
            PR1-1 gene is organized into two exons (274 and 439 bp in size),
            interrupted by a 164 bp intron.
FEATURES             Location/Qualifiers
     source          1. .930
                     /organism="Petroselinum crispum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4043"
                     /clone="lambda ps11"
                     /clone_lib="lambda EMBL4."
     precursor_RNA   841. .>930
                     /note="PR1-1 transcript"
     mRNA            841. .>930
                     /note="PR1-1 mRNA exon 1"
     CDS             927. .>930
                     /note="unnamed protein product; Protein sequence is in
                     conflict with the conceptual translation; PR1-1"
                     /codon_start=1
                     /protein_id="CAA31084.1"
                     /db_xref="GI:4379347"
                     /translation="Y"
ORIGIN

  Query Match           100.0%;  Score 30;  DB 8;  Length 930;
  Best Local Similarity   100.0%;  Pred. No. 0.0021;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATTCGCC 30
              ||||||||||||||||||||||||||||||

Db        589 CACACTTAATTTGACCGAGTAACATTCGCC 618


RESULT 3
PCU48862
LOCUS       PCU48862                  2011 bp    DNA     linear   PLN 12-FEB-1997
DEFINITION  Petroselinum crispum pathogenesis-related protein (pr1-1) gene,
            complete cds.
ACCESSION   U48862
VERSION     U48862.1  GI:1840040
KEYWORDS    .
SOURCE      Petroselinum crispum (parsley)
  ORGANISM  Petroselinum crispum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Apiales; Apiaceae; Apioideae; apioid
            superclade; Apium clade; Petroselinum.
REFERENCE   1  (bases 1 to 2011)
  AUTHORS   Rushton,P.J., Torres,J.T., Parniske,M., Wernert,P., Hahlbrock,K.
            and Somssich,I.E.
  TITLE     Interaction of elicitor-induced DNA-binding proteins with elicitor
            response elements in the promoters of parsley PR1 genes
  JOURNAL   EMBO J. 15 (20), 5690-5700 (1996)
  MEDLINE   97051827
   PUBMED   8896462
REFERENCE   2  (bases 1 to 2011)
  AUTHORS   Somssich,I.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-1996) Imre E. Somssich, Dept. of Biochemistry,
            Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne Weg
            10, Koeln, 50829, Germany
FEATURES             Location/Qualifiers
     source          1. .2011
                     /organism="Petroselinum crispum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4043"
     gene            113. .1727
                     /gene="pr1-1"
     CAAT_signal     113. .134
                     /gene="pr1-1"
                     /note="elicitor-response element 2"
     misc_feature    589. .616
                     /gene="pr1-1"
                     /note="elicitor-response element 1"
     misc_feature    695. .716
                     /gene="pr1-1"
                     /note="constitutive in vivo DNA footprint region"
     CAAT_signal     739. .743
                     /gene="pr1-1"
     CAAT_signal     773. .777
                     /gene="pr1-1"
     TATA_signal     814. .820
                     /gene="pr1-1"
     mRNA            join(841. .1114,1279. .1761)
     prim_transcript 841. .1761
     5'UTR           841. .930
                     /gene="pr1-1"
     CDS             join(931. .1114,1279. .1562)
                     /gene="pr1-1"
                     /function="unknown"
                     /note="PR1-1; belongs to the PR-10 class of intracellular
                     PR proteins"
                     /codon_start=1
                     /product="pathogenesis-related protein 1"
                     /protein_id="AAB47234.1"
                     /db_xref="GI:1840041"
                     /translation="MGVQKSEVETTSSVSAEKLFKGLCLDIDTLLPQVLPGAIKSSET
                     LEGDGGVGTVKLVHLGDASPFKTMKQKVDAIDKATFTYSYSIIDGDILLGFIESINNH
                     FTAVPNADGGCTVKSTIIFNTKGDAVVPEENIKFANDQNLTIFKAVEAYLIAN"
     3'UTR           1563. .1761
     polyA_signal    1720. .1727
                     /gene="pr1-1"
ORIGIN
```

```
  Query Match             100.0%;  Score 30;  DB 8;  Length 2011;
  Best Local Similarity   100.0%;  Pred. No. 0.0021;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATTCGCC 30
              ||||||||||||||||||||||||||||||
Db        589 CACACTTAATTTGACCGAGTAACATTCGCC 618


RESULT 4
PCU48863
LOCUS       PCU48863               1312 bp    DNA     linear   PLN 12-FEB-1997
DEFINITION  Petroselinum crispum pathogenesis-related protein (pr1-2) gene,
            promoter region and most of exon 1.
ACCESSION   U48863
VERSION     U48863.1  GI:1840042
KEYWORDS    .
SOURCE      Petroselinum crispum (parsley)
  ORGANISM  Petroselinum crispum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Apiales; Apiaceae; Apioideae; apioid
            superclade; Apium clade; Petroselinum.
REFERENCE   1  (bases 1 to 1312)
  AUTHORS   Rushton,P.J., Torres,J.T., Parniske,M., Wernert,P., Hahlbrock,K.
            and Somssich,I.E.
  TITLE     Interaction of elicitor-induced DNA-binding proteins with elicitor
            response elements in the promoters of parsley PR1 genes
  JOURNAL   EMBO J. 15 (20), 5690-5700 (1996)
  MEDLINE   97051827
   PUBMED   8896462
REFERENCE   2  (bases 1 to 1312)
  AUTHORS   Somssich,I.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-1996) Imre E. Somssich, Dept. of Biochemistry,
            Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne Weg
            10, Koeln, 50829, Germany
FEATURES             Location/Qualifiers
     source          1. .1312
                     /organism="Petroselinum crispum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4043"
     gene            1. .1312
                     /gene="pr1-2"
     promoter        1. .1082
                     /gene="pr1-2"
     misc_feature    670. .703
                     /gene="pr1-2"
                     /note="elicitor-response element 3"
     misc_feature    814. .858
                     /gene="pr1-2"
                     /note="elicitor-response element 1"
     misc_feature    929. .950
                     /gene="pr1-2"
                     /note="constitutive in vivo footprinted region"
     CAAT_signal     969. .973
                     /gene="pr1-2"
     CAAT_signal     1003. .1007
                     /gene="pr1-2"
     TATA_signal     1056. .1062
                     /gene="pr1-2"
     mRNA            1083. .>1312
                     /gene="pr1-2"
     5'UTR           1083. .1172
                     /gene="pr1-2"
     CDS             1173. .>1312
                     /gene="pr1-2"
                     /function="unknown"
                     /note="PR1-2; belongs to the PR-10 class of intracellular
                     PR proteins"
                     /codon_start=1
                     /product="pathogenesis-related protein 1-2"
                     /protein_id="AAB47235.1"
                     /db_xref="GI:1840043"
                     /translation="MGVQKSEVETTSSVSAEKLFKGLCLDIDTLLPQVLPGAIKSSET
                     LEG"
ORIGIN

  Query Match             94.7%;  Score 28.4;  DB 8;  Length 1312;
  Best Local Similarity   96.7%;  Pred. No. 0.012;
  Matches   29;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATTCGCC 30
              |||||||||||||||||||||||||| ||
Db        816 CACACTTAATTTGACCGAGTAACATTCACC 845


RESULT 5
AL670670/c
LOCUS       AL670670               91811 bp    DNA     linear   ROD 23-MAY-2002
DEFINITION  Mouse DNA sequence from clone RP23-234O7 on chromosome X, complete
            sequence.
ACCESSION   AL670670
VERSION     AL670670.7  GI:21213385
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 91811)
  AUTHORS   Clark,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 25, 2002 this sequence version replaced gi:21065347.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-234O7 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .91811
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="X"
                     /clone="RP23-234O7"
                     /clone_lib="RPCI-23"
ORIGIN

  Query Match             68.0%;  Score 20.4;  DB 10;  Length 91811;
  Best Local Similarity   80.0%;  Pred. No. 92;
  Matches   24;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATTCGCC 30
              ||||||| || |||| ||| || ||||| |
Db      37196 CACACTTTATGTGACTGAGAAAGATTCGAC 37167
```

```
RESULT 6
AL513220/c
LOCUS       AL513220               94638 bp    DNA     linear   PRI 17-APR-2002
DEFINITION  Human DNA sequence from clone RP11-566C13 on chromosome 1, complete
            sequence.
ACCESSION   AL513220
VERSION     AL513220.9  GI:20196551
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Matthews,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 18, 2002 this sequence version replaced gi:19699427.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-566C13 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .94638
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-566C13"
                     /clone_lib="RPCI-11.2"
ORIGIN

  Query Match             68.0%;  Score 20.4;  DB 9;  Length 94638;
  Best Local Similarity   80.0%;  Pred. No. 92;
  Matches   24;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATTCGCC 30
              ||||||| |||| ||| ||| ||||| | |
Db       1781 CACACTTGATTTTACCCAGTGACATTAGAC 1752


RESULT 7
HS423B22/c
LOCUS       HS423B22               112224 bp    DNA     linear   PRI 04-APR-2001
DEFINITION  Human DNA sequence from clone RP3-423B22 on chromosome 1p33-35.3
            Contains part of a gene for a novel protein similar to KIAA0323,
            KIAA0615 and C.elegans C30F12.1, a ferritin H pseudogene, the gene
            for RPS27 (ribosomal protein S27  (metallopanstimulin 1)), the gene
            for axonemal dynein light chain (hp28), the gene for a novel
            nucleolar guanosine 5'-triphosphate binding protein autoantigen,
            STSs, GSSs and a CpG Island, complete sequence.
ACCESSION   AL034379
VERSION     AL034379.8  GI:5918013
KEYWORDS    HTG; autoantigen; CpG Island; dynein; ferritin; hp28; KIAA0323;
            KIAA0615; metallopanstimulin; ribosomal protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 112224)
  AUTHORS   Pearce,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 22, 1999 this sequence version replaced gi:5870406.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP3-423B22 is from the library RPCI-3 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP3-423B22.
FEATURES             Location/Qualifiers
     source          1. .112224
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p33-35.3"
                     /clone="RP3-423B22"
                     /clone_lib="RPCI-3"
     misc_feature    885. .1249
                     /note="match: GSS: Em:AQ188117"
     misc_feature    complement(1836. .2220)
                     /note="match: GSS: Em:AQ124662"
     gene            join(3734. .3873,5045. .5279,5878. .5984)
                     /gene="dJ423B22.1"
     CDS             join(<3734. .3873,5045. .5279,5878. .>5984)
                     /gene="dJ423B22.1"
                     /note="match: ESTs: Em:AI832215
                     match: proteins: Tr:O15037 Tr:O75113 Tr:O95530 Wp:CE16881"
                     /codon_start=2
                     /evidence=not_experimental
                     /product="dJ423B22.1 (novel protein similar to KIAA0323,
                     KIAA0615 and C.elegans C30F12.1)"
                     /protein_id="CAC36083.1"
                     /db_xref="GI:13559173"
                     /db_xref="SPTREMBL:Q9BQZ9"
                     /translation="HGNKEVFSCRGILLAVNWFLERGHTDITVFVPSWRKEQPRPDVP
                     ITDQHILRELEKKKILVFTPSRRVGGKRVVCYDDRFIVKLAYESDGIVVSNDTYRDLQ
```

```
                GERQEWKRFIEERLLMYSFVNDKFMPPDDPLGRHGPSLDNFLRKKPLTLEHRKQPCPY
                "
repeat_region   16005. .16034
                /note="15 copies 2 mer aa 86% conserved"
misc_feature    16038. .16217
                /note="match: STS: Em:G25594 Em:G27659"
misc_feature    16758. .17140
                /note="match: STS: Em:G38585"
misc_feature    16768. .17099
                /note="match: STS: Em:G29990"
misc_feature    complement(16958. .17522)
                /note="match: GSS: Em:AF046586"
repeat_region   17051. .17124
                /note="37 copies 2 mer aa 71% conserved"
misc_feature    complement(17893. .18413)
                /note="match: GSS: Em:AQ635919"
misc_feature    complement(17988. .18405)
                /note="match: GSS: Em:AQ220399"
misc_feature    complement(21756. .22258)
                /note="match: GSS: Em:AQ804436"
misc_feature    22259. .22501
                /note="match: GSS: Em:AQ312816"
misc_feature    22259. .22496
                /note="match: GSS: Em:AQ277563"
misc_feature    22272. .22698
                /note="match: GSS: Em:AQ880377"
repeat_region   25551. .25607
                /note="3 copies 19 mer 98% conserved"
repeat_region   25981. .26022
                /note="21 copies 2 mer tt 78% conserved"
repeat_region   28245. .28270
                /note="13 copies 2 mer ac 100% conserved"
misc_feature    complement(28745. .29200)
                /note="match: GSS: Em:AQ279885"
repeat_region   37708. .37903
                /note="98 copies 2 mer cc 59% conserved"
misc_feature    37745. .38622
                /note="CpG island"
                /evidence=not_experimental
repeat_region   39158. .39335
                /note="L1M3b repeat: matches 1519. .1695 of consensus"
repeat_region   40696. .40874
                /note="MER51-internal repeat: matches 2762. .2946 of
                consensus"
repeat_region   40913. .41325
                /note="MER51-internal repeat: matches 2340. .2762 of
                consensus"
repeat_region   41425. .41512
                /note="MSTB-internal repeat: matches 53. .143 of consensus"
repeat_region   41513. .42266
                /note="pTR5 repeat: matches 743. .1476 of consensus"
repeat_region   43004. .43054
                /note="MSTB-internal repeat: matches 6. .59 of consensus"
repeat_region   44443. .44578
                /note="L1M3b repeat: matches 1234. .1410 of consensus"
repeat_region   48476. .48626
                /note="MER83-internal repeat: matches 1652. .1790 of
                consensus"
repeat_region   49239. .49303
                /note="MER83-internal repeat: matches 1790. .1857 of
                consensus"
repeat_region   49659. .49763
                /note="MER83-internal repeat: matches 2718. .2814 of
                consensus"
misc_feature    complement(52297. .52760)
                /note="match: GSS: Em:AQ474424"
misc_feature    complement(57547. .57900)
                /note="match: GSS: Em:AQ296347"
misc_feature    complement(59824. .60367)
                /note="match: GSS: Em:AQ547506"
misc_feature    60547. .60754
                /note="match: GSS: Em:AQ473963"
misc_feature    60701. .60797
                /note="match: STS: Em:HSPE43A11 Em:HSPE49C01"
gene            complement(join(61207. .61456,63601. .63972))
                /gene="dJ423B22.2"
CDS             complement(join(<61207. .61456,63601. .>63972))
                /gene="dJ423B22.2"
                /note="match: proteins: Wp:CE08529"
                /codon_start=2
                /evidence=not_experimental
                /product="dJ423B22.2 (novel protein similar to C.elegans
                protein CE08529)"
                /protein_id="CAC36084.1"
                /db_xref="GI:13559174"
                /db_xref="SPTREMBL:Q9BQZ8"
                /translation="RREHRQRNDVGGGGSESQELVPRPGGNNKEKEVPAKEKPSFELS
                GALLEDTNTFRGVVIKYSEPPEARIPKKRWRLYPFKNDEVLPVMYIHRQSAYLLGRHR
                RIADIPIDHPSCSKQHAVFQYRLVEYTRADGTVGRRVKPYIIDLGSGNGTFLNNKRIE
                PQRYYELKEKDVLKFGFSSREYVLLHESSDTSEIDRKDDEDEEEEEE"
repeat_region   65164. .65853
                /note="MER44C repeat: matches 1. .728 of consensus"
misc_feature    68077. .68389
                /note="match: GSS: Em:AQ839815"
gene            68207. .68745
                /gene="dJ423B22.3"
                /pseudo
CDS             68207. .68745
                /gene="dJ423B22.3"
                /note="dJ423B22.3 (ferritin H pseudogene)
                match: cDNAs: Em:L20941 Em:M97164 Em:E08663 Em:D86627
                Em:M12937 Em:M63912 Em:M99692 Em:D15071 Em:J04755
                Em:M11146 Em:S77386 Em:U54800 Em:X00318 Em:AB003093
                match: ESTs: Em:AI816415 Em:AI346347 Em:AI460154
                Em:AA632163 Em:AA775128 Em:AI092944 Em:AA057837
                Em:AI929679 Em:AI935207 Em:AA642200 Em:AI345942
                Em:AA010815 Em:AI346840 Em:AI582308 Em:AA770636
                Em:AI246580 Em:AA131039 Em:AA604610 Em:AI079884 Em:T57843
                Em:AA056594 Em:AI830614 Em:N78958 Em:AI879024 Em:AI749039
                Em:AI864456 Em:AA161102 Em:C18478 Em:AI879269 Em:AA165061
                Em:AW021110 Em:AA963209 Em:AA725087 Em:AA157187
                Em:AA083658 Em:D51220 Em:AA186730 Em:AI950656 Em:AA122004
                Em:N52036 Em:H66415 Em:AW046761 Em:AA086011 Em:AA456170
                Em:AA565350 Em:AA084238 Em:AI247121 Em:AA935516
                Em:AI075405 Em:AA657366 Em:AA151760 Em:AI815565
                Em:AI785093 Em:AI929672 Em:N70862 Em:AI939639 Em:AA102093
                Em:AA112615 Em:AA701506 Em:W31508 Em:AA831015 Em:AA044647
                Em:AA602089 Em:AI929641 Em:AA052988 Em:AA482667 Em:Z99431
                Em:AI831973 Em:C17972 Em:AA829459 Em:C18189 Em:AA225120
                Em:AA296646 Em:AA662010 Em:AA191015 Em:AA303930 Em:R67061
                Em:AA188320 Em:AI373411 Em:AA371395 Em:AA918127 Em:N89764
                Em:AA057849 Em:D54420 Em:AA808377 Em:AI876587 Em:AA132507

  Query Match             68.0%;  Score 20.4;  DB 9;  Length 112224;
  Best Local Similarity   80.0%;  Pred. No. 92;
  Matches   24;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATTCGCC 30
              ||||||| |||| ||| ||| ||||| | |
Db     112005 CACACTTGATTTTACCCAGTGACATTAGAC 111976


RESULT 8
AL355880/c
LOCUS       AL355880               153023 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-195A8, 4 unordered pieces.
ACCESSION   AL355880 AC027731
VERSION     AL355880.7  GI:11876042
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Wallis,J.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On or before May 15, 2001 this sequence version replaced
            gi:8077031, gi:11544491.
            Draft Sequence Produced by Whitehead Institute/MIT Center for
            Genome Research, 320 Charles Street,
            Cambridge, MA 02141, USA
            http://www-seq.wi.mit.edu
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA195A8
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: M13; M77815; 34% of reads
            Sequencing vector: plasmid; L08752; 65% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 150796 bases at least Q40
            Consensus quality: 151533 bases at least Q30
            Consensus quality: 151958 bases at least Q20
            Insert size: 152723; sum-of-contigs
            Insert size: 155229; 2.8% error; agarose-fp
            Quality coverage: 9.72x in Q20 bases; sum-of-contigs Quality
            coverage: 9.97x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   144938: contig of 144938 bp in length
            *  144939   145038: gap of 100 bp
            *  145039   147529: contig of 2491 bp in length
            *  147530   147629: gap of 100 bp
            *  147630   149636: contig of 2007 bp in length
            *  149637   149736: gap of 100 bp
            *  149737   153023: contig of 3287 bp in length.
FEATURES             Location/Qualifiers
     source          1. .153023
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-195A8"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .144938
                     /note="assembly_fragment:00424
                     clone_end:T7
                     vector_side:left"
     misc_feature    145039. .147529
                     /note="assembly_fragment:00960"
     misc_feature    147630. .149636
                     /note="assembly_fragment:01053"
     misc_feature    149737. .153023
                     /note="assembly_fragment:03302"
ORIGIN

  Query Match             68.0%;  Score 20.4;  DB 2;  Length 153023;
  Best Local Similarity   80.0%;  Pred. No. 92;
  Matches   24;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATTCGCC 30
              ||||||| |||| ||| ||| ||||| | |
Db     141556 CACACTTGATTTTACCCAGTGACATTAGAC 141527
```

```
RESULT 9
BX571867
LOCUS       BX571867              336609 bp    DNA     linear   BCT 26-SEP-2003
DEFINITION  Photorhabdus luminescens subsp. laumondii TTO1 complete genome;
            segment 9/17.
ACCESSION   BX571867 BX470251
VERSION     BX571867.1  GI:36785705
KEYWORDS    complete genome.
SOURCE      Photorhabdus luminescens subsp. laumondii TTO1
  ORGANISM  Photorhabdus luminescens subsp. laumondii TTO1
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Photorhabdus.
REFERENCE   1
  AUTHORS   Duchaud,E., Rusniok,C., Frangeul,L., Buchrieser,C., Taourit,S.,
            Bocs,S., Boursaux-Eude,C., Chandler,M., Dassa,E., Derose,R.,
            Derzelle,S., Freyssinet,G., Gaudriault,S., Givaudan,A., Glaser,P.,
            Medigue,C., Lanois,A., Powell,K., Siguier,P., Wingate,V.,
            Zouine,M., Boemare,N., Danchin,A. and Kunst,F.
  TITLE     Complete genome sequence of the entomapathogenic bacterium
            Photorhabdus luminescens
  JOURNAL   Nat. Biotechnol. 11 (1) (2003) In press
REFERENCE   2
  AUTHORS   Duchaud,E., Frangeul,L., Rusniok,C. and Kunst,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2003) L. Frangeul, Institut Pasteur, Genopole, 25
            rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail:
            lfrangeu@pasteur.fr, fkunst@pasteur.fr
FEATURES             Location/Qualifiers
     source          1. .336609
                     /organism="Photorhabdus luminescens subsp. laumondii TTO1"
                     /mol_type="genomic DNA"
                     /strain="TTO1"
                     /db_xref="taxon:243265"
     gene            complement(425. .1409)
                     /locus_tag="plu2374"
     CDS             complement(425. .1396)
                     /locus_tag="plu2374"
                     /function="Unknown"
                     /note="unnamed protein product; Highly similar to putative
                     cytochrome oxidase, YfeH protein of Escherichia coli"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE14684.1"
                     /db_xref="GI:36785706"
                     /translation="MTWWQKLKIDSFLMILVTVVVIATFFPCEGEIKVWFQRLTTAAI
                     ALLFFMHGAKLSRQAIIAGMAHWRLHLVVFLSTFVLFPIIGLGLNLLVPEIMTSTVYW
                     GFLYLCALPATVQSAIAFTSVAGGNVAAAICSASASSILGVFLSPILVGLMIHTQGEG
                     SADTLEAIKAIILQLMVPFIIGHLSRPLIARWVEKHKKLLNMTDRSSILLVVYVAFSE
                     AVVEGIWHKIDGYSLLTIGVVCCVLLAVVLVINTYIARWLGFNKEDEITIVFCGSKKS
                     LANGVPMANVLFPAATVGVMLLPLMIFHQIQLMVCAVLAQRYAKRLK"
     RBS             complement(1404. .1409)
                     /locus_tag="plu2374"
     gene            complement(1605. .2653)
                     /locus_tag="plu2375"
     CDS             complement(1605. .2642)
                     /locus_tag="plu2375"
                     /function="Unknown"
                     /note="unnamed protein product; Similar to hypothetical
                     oxidoreductase YdgJ of Escherichia coli"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE14688.1"
                     /db_xref="GI:36785707"
                     /translation="MRELLKVGLVGYGYASKTFHAPLIAGTAGVKLAAISSSDAEKVK
                     SDWPNIPVVSRPEDLFNDPDIDLIVIPTPNDTHYPLAQKALAAGKHVVVDKPFTITVE
                     QAESLKQQAEKADLLLSVFHNRRWDSGFLTIKSLIKENRLGKLKYFASHFDRYRPIIR
                     QRWREAAGQGSGIWYDLGPHLLDQAIQLFGKPQSITADLGMIRPQAEAVDYFHVLLNY
                     PDIKVVLHATTLAAAESAVYTLHGMAGSYVKYGLDPQEERLKAGERPPHADWGYDIRD
                     GIVTLSHGEELVTKHVPTLPGNYSAYYAAIRDAIMLGKPNPVTANEAIQIMKLIEAGI
                     QSAEQQRTVAI"
     RBS             complement(2648. .2653)
                     /locus_tag="plu2375"
     gene            2826. .3266
```

```
                     /locus_tag="plu2376"
     CDS             2826. .3266
                     /locus_tag="plu2376"
                     /function="Unknown"
                     /note="unnamed protein product; Similar to probable
                     membrane protein YdgK of Escherichia coli"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE14690.1"
                     /db_xref="GI:36785708"
                     /translation="MYQTNDYHRINGWVLAPAAYLIMTFLSASVMLALYIMTFFSQNN
                     MLSAGNNHFTLMWFLSVAITAVMWCFTLWVLKLLFIRSKKFPHIFIIWLIVTVLIALK
                     TFAFSPIPDQMALRALAWPLLAAAIFVPYIKRSHRVKMTFTQDR"
     gene            3393. .3974
                     /gene="rnfA"
                     /locus_tag="plu2377"
     CDS             3393. .3974
                     /gene="rnfA"
                     /locus_tag="plu2377"
                     /codon_start=1
                     /transl_table=11
                     /product="Electron transport complex protein RnfA"
                     /protein_id="CAE14692.1"
                     /db_xref="GI:36785709"
                     /translation="MTDYFLLFVGTVLVNNFVLVKFLGLCPFMGVSKKLETAIGMGFA
                     TTFVMTLASVCSWLVNTFILLPLDLIYLRTLSFILVIAVVVQFTELVVRKTSPTLYRL
                     LGIFLPLITTNCAVLGVALLNINQSHDFLQSAIYGFGAAAGFSLVMVLFAAIRERLAV
                     ADVPAPFKGSSIGLITAGLMSLAFMGFSGLVKF"
     gene            3974. .4606
                     /gene="rnfB"
                     /locus_tag="plu2378"
     CDS             3974. .4606
                     /gene="rnfB"
                     /locus_tag="plu2378"
                     /codon_start=1
                     /transl_table=11
                     /product="Electron transport complex protein RnfB"
                     /protein_id="CAE14694.1"
                     /db_xref="GI:36785710"
                     /translation="MTSLWIAIGALSALGLIFGLILGFAARRFQVEEDPIVENIDDIL
                     PQSQCGQCGYPGCRPYAEAVANNGEMINKCAPGGEQVMLKISELLGIDPQPLDGDANL
                     QNPARKVAFIDEENCIGCTKCIQACPVDAIVGATRAMHTIVEDLCTGCDLCVAPCPTD
                     CITMIPVATTTSNWKWDLKTIPVRNIPVQPSHAISIKSIPIKNIEAEHHV"
     gene            4587. .4592
                     /gene="rnfC"
                     /locus_tag="plu2379"
     RBS             4587. .4592
                     /gene="rnfC"
                     /locus_tag="plu2379"
     gene            4599. .7157
                     /gene="rnfC"
                     /locus_tag="plu2379"
     CDS             4599. .7157
                     /gene="rnfC"
                     /locus_tag="plu2379"
                     /codon_start=1
                     /transl_table=11
                     /product="Electron transport complex protein RnfC"
                     /protein_id="CAE14696.1"
                     /db_xref="GI:36785711"
                     /translation="MFNLFNLLNKNKIWDFAGGIHPPEMKLQSSQSPLRITPLPEELI
                     IPLQQHLGTEGELLVKVGDEVLKGQPLTFGTGRTVPVHASTSGTIIAIEPCITAHPSG
                     LKELCVRLRPDGKDQWGERIQTPDYQSLDADTILNRIHQAGIAGLGGAGFPTAAKLQS
                     GRHGLKTLIINAAECEPYITADDRLMQEHANELIEGIRILEHLLKPEQILIGIEDNKP
                     AAIKKLKTVLKGDKSIIVRVIPTKYPSGGAKQLTKILTGKEVPSGGRSSTIGVLMQNV
                     GTVVAIKRAIIDGEPLIERVVTITGEAVTSPGNFWARLGTPVQFLLQLAGFAPQSEQM
                     VIMGGPLMGFTLPDLNVPIVKISNCILAPSHQEMDSNTTEEACIRCGLCVEACPARLL
                     PQQLYWFSRGQEHEKARDHNLFDCIECGACAYVCPSNIPLVQYYRQEKAEIRALDQEN
                     RRSTEAKIRFEAKQARMEREKLARDERHKKSAVQVDNKDKSAVQDALSRVKEKQANAG
                     KVVAIEPGQLPDNAAVIAARQARKEQLRARQAEKQAVENMSITPSEPTSAEDPRKAAL
                     AAAIARAKAKKAAQNQSETPSPEAREPETTTSEKQDLHKDAVAAAIARVKAKKAQTQP
                     LTETQAETPPAEETDPRKAAVAAAIARVKAKKAQTQPLTETQAETPPAEETDPRKAAV
                     AAAIARVKAKKAQTQPLTETQAETPPAEETDPRKAAVAAAIARVKAKKAQTQQLTETQ
                     AETPPAEETDPRKAAVAAAIARVKAKKAQTQQLTETQAETPPAEETDPRKAAVAAAIA
                     RVKAKKAQTEQLAETQVETPPAEETDPRKAAVAAAIARVKAKKDAQEQQKISIE"
     gene            7170. .8249
                     /gene="rfnD"
                     /locus_tag="plu2380"
     CDS             7170. .8249
                     /gene="rfnD"
                     /locus_tag="plu2380"
                     /codon_start=1
                     /transl_table=11
                     /product="Electron transport complex protein RnfD"
                     /protein_id="CAE14698.1"
                     /db_xref="GI:36785712"
                     /translation="MKFRPVQTDNKKLKVASSPFTHNQQSTSRIMLWVVLAAIPGIAV
                     QTYFPSYGTLFQLLLAVITALLAESFALSLRKQAIMPRLQDNSALLTGLLLGISLPPL
                     APWWLIVLGTAFAVIIAKHLYGGLGQNPFNPAMVGYVVLLISFPVQMTGWLPPMDLLA
                     TQITPLDSLMVIFTGQTPDGTTLQQLQQGFDGISQATPLDSFKTGLLTHHIDDVLHQP
                     ILQGSLAGIGWQWVNIAYLVGGLILLNRKVITWQIPVAFILVLGLCALTSWLIDPSRY
                     SPPLLQLFSGATMLGAFFIATDPVSASTTPKGRLVFGAMIGLLVWVIRVYGGYPDAIA
                     FAVLLANITVPLIDHYTQPRVYGHK"
     gene            join(8249. .8254,8259. .8888)
                     /gene="rfnG"
                     /locus_tag="plu2381"
     RBS             8249. .8254
                     /gene="rfnG"
                     /locus_tag="plu2381"
     CDS             8259. .8888
                     /gene="rfnG"
                     /locus_tag="plu2381"
                     /codon_start=1
                     /transl_table=11
                     /product="Electron transport complex protein RnfG"
                     /protein_id="CAE14701.1"
                     /db_xref="GI:36785713"
                     /translation="MLETMRRHGITLAIFAALTTGLTAVVNSLTKSTIAEQAALQHKS
                     LLDQVIPPALYDNDIQNECYLVNANALGNNLPHRLYLARKNGHPVAAALESTAPDGYS
                     GAIQLLVGADFSGKVLGVRVTEHHETPGLGDKIETRISNWINVFSGKTITSNHDQHWA
                     VKKDGGEFDQFTGATITPRAVVNSVKRTALYLQTIPEHLSSLESCGEKS"
     gene            8873. .8878
                     /gene="rfnE"
                     /locus_tag="plu2382"
     RBS             8873. .8878
                     /gene="rfnE"
                     /locus_tag="plu2382"
     gene            8885. .9583
                     /gene="rfnE"
                     /locus_tag="plu2382"
     CDS             8885. .9583
                     /gene="rfnE"
                     /locus_tag="plu2382"
                     /codon_start=1
                     /transl_table=11
                     /product="Electron transport complex protein RnfE"
                     /protein_id="CAE14703.1"
                     /db_xref="GI:36785714"
                     /translation="MSETKNLFIQGLWKNNSALVQLLGLCPLLAVSSTATNALGLGLA

  Query Match             66.0%;  Score 19.8;  DB 1;  Length 336609;
  Best Local Similarity   91.3%;  Pred. No. 1.8e+02;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          5 CTTAATTTGACCGAGTAACATTC 27
              ||||| || ||||||||||||||
Db      12713 CTTAAATTCACCGAGTAACATTC 12735


RESULT 10
AX770906/c
LOCUS       AX770906               349980 bp    DNA     linear   PAT 02-JUL-2003
DEFINITION  Sequence 37 from Patent WO02094867.
ACCESSION   AX770906
VERSION     AX770906.1  GI:32438070
KEYWORDS    .
```

```
SOURCE      Photorhabdus luminescens
  ORGANISM  Photorhabdus luminescens
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Photorhabdus.
REFERENCE   1
  AUTHORS   Duchaud,E., Taourit,S., Glaser,P., Frangeul,L., Kunst,F.,
            Danchin,A. and Buchrieser,C.
  TITLE     Sequence of the Photorhabdus luminescens strain TT01 genome and
            uses
  JOURNAL   Patent: WO 02094867-A 37 28-NOV-2002;
            INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
            SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="Photorhabdus luminescens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:29488"
                     /note="sequence length too big . Splitted into 2
                     sequences.~seq 00037: 1 to 349980~seq 10786: 300001 to
                     530312"
ORIGIN

  Query Match             66.0%;  Score 19.8;  DB 6;  Length 349980;
  Best Local Similarity   91.3%;  Pred. No. 1.8e+02;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          5 CTTAATTTGACCGAGTAACATTC 27
              ||||| || ||||||||||||||
Db     110745 CTTAAATTCACCGAGTAACATTC 110723


RESULT 11
AC100379
LOCUS       AC100379               59008 bp    DNA     linear   HTG 28-NOV-2002
DEFINITION  Mus musculus clone RP23-131C19, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC100379
VERSION     AC100379.3  GI:25815482
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 59008)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-131C19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 59008)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 59008)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 28, 2002 this sequence version replaced gi:24962776.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L15213
                Center clone name: 131_C_19
            --------------.
            * NOTE: This record contains 74 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      724: contig of 724 bp in length
            *      725      824: gap of 100 bp
            *      825     1527: contig of 703 bp in length
            *     1528     1627: gap of 100 bp
            *     1628     2324: contig of 697 bp in length
            *     2325     2424: gap of 100 bp
            *     2425     3114: contig of 690 bp in length
            *     3115     3214: gap of 100 bp
            *     3215     3925: contig of 711 bp in length
            *     3926     4025: gap of 100 bp
            *     4026     4728: contig of 703 bp in length
            *     4729     4828: gap of 100 bp
            *     4829     5538: contig of 710 bp in length
            *     5539     5638: gap of 100 bp
            *     5639     6354: contig of 716 bp in length
            *     6355     6454: gap of 100 bp
            *     6455     7158: contig of 704 bp in length
            *     7159     7258: gap of 100 bp
            *     7259     7977: contig of 719 bp in length
            *     7978     8077: gap of 100 bp
            *     8078     8741: contig of 664 bp in length
            *     8742     8841: gap of 100 bp
            *     8842     9527: contig of 686 bp in length
            *     9528     9627: gap of 100 bp
            *     9628    10313: contig of 686 bp in length
            *    10314    10413: gap of 100 bp
            *    10414    11099: contig of 686 bp in length
            *    11100    11199: gap of 100 bp
```

```
*    11200    11915: contig of 716 bp in length
*    11916    12015: gap of 100 bp
*    12016    12717: contig of 702 bp in length
*    12718    12817: gap of 100 bp
*    12818    13522: contig of 705 bp in length
*    13523    13622: gap of 100 bp
*    13623    14304: contig of 682 bp in length
*    14305    14404: gap of 100 bp
*    14405    15085: contig of 681 bp in length
*    15086    15185: gap of 100 bp
*    15186    15892: contig of 707 bp in length
*    15893    15992: gap of 100 bp
*    15993    16685: contig of 693 bp in length
*    16686    16785: gap of 100 bp
*    16786    17467: contig of 682 bp in length
*    17468    17567: gap of 100 bp
*    17568    18269: contig of 702 bp in length
*    18270    18369: gap of 100 bp
*    18370    19092: contig of 723 bp in length
*    19093    19192: gap of 100 bp
*    19193    19917: contig of 725 bp in length
*    19918    20017: gap of 100 bp
*    20018    20719: contig of 702 bp in length
*    20720    20819: gap of 100 bp
*    20820    21521: contig of 702 bp in length
*    21522    21621: gap of 100 bp
*    21622    22295: contig of 674 bp in length
*    22296    22395: gap of 100 bp
*    22396    23087: contig of 692 bp in length
*    23088    23187: gap of 100 bp
*    23188    23889: contig of 702 bp in length
*    23890    23989: gap of 100 bp
*    23990    24708: contig of 719 bp in length
*    24709    24808: gap of 100 bp
*    24809    25427: contig of 619 bp in length
*    25428    25527: gap of 100 bp
*    25528    26230: contig of 703 bp in length
*    26231    26330: gap of 100 bp
*    26331    27004: contig of 674 bp in length
*    27005    27104: gap of 100 bp
*    27105    27800: contig of 696 bp in length
*    27801    27900: gap of 100 bp
*    27901    28595: contig of 695 bp in length
*    28596    28695: gap of 100 bp
*    28696    29354: contig of 659 bp in length
*    29355    29454: gap of 100 bp
*    29455    30162: contig of 708 bp in length
*    30163    30262: gap of 100 bp
*    30263    30953: contig of 691 bp in length
*    30954    31053: gap of 100 bp
*    31054    31766: contig of 713 bp in length
*    31767    31866: gap of 100 bp
*    31867    32567: contig of 701 bp in length
*    32568    32667: gap of 100 bp
*    32668    33377: contig of 710 bp in length
*    33378    33477: gap of 100 bp
*    33478    34200: contig of 723 bp in length
*    34201    34300: gap of 100 bp
*    34301    35022: contig of 722 bp in length
*    35023    35122: gap of 100 bp
*    35123    35765: contig of 643 bp in length
*    35766    35865: gap of 100 bp
*    35866    36564: contig of 699 bp in length
*    36565    36664: gap of 100 bp
*    36665    37371: contig of 707 bp in length
*    37372    37471: gap of 100 bp
*    37472    38173: contig of 702 bp in length
*    38174    38273: gap of 100 bp
*    38274    38967: contig of 694 bp in length
*    38968    39067: gap of 100 bp
*    39068    39772: contig of 705 bp in length
*    39773    39872: gap of 100 bp
*    39873    40562: contig of 690 bp in length
*    40563    40662: gap of 100 bp
*    40663    41372: contig of 710 bp in length
*    41373    41472: gap of 100 bp
*    41473    42181: contig of 709 bp in length
*    42182    42281: gap of 100 bp
*    42282    42999: contig of 718 bp in length
*    43000    43099: gap of 100 bp
*    43100    43805: contig of 706 bp in length
*    43806    43905: gap of 100 bp
*    43906    44612: contig of 707 bp in length
*    44613    44712: gap of 100 bp
*    44713    45407: contig of 695 bp in length
*    45408    45507: gap of 100 bp
*    45508    46205: contig of 698 bp in length

  Query Match             65.3%;  Score 19.6;  DB 2;  Length 59008;
  Best Local Similarity   84.6%;  Pred. No. 2.3e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATT 26
              ||||||| |||||| | ||||| |||
Db       9942 CACACTTCATTTGAACTAGTAAAATT 9967


RESULT 12
AC100388/c
LOCUS       AC100388               60978 bp    DNA     linear   HTG 22-NOV-2001
DEFINITION  Mus musculus clone RP23-133A3, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC100388
VERSION     AC100388.1  GI:17047754
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 60978)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-133A3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 60978)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
```

```
-------------- Project Information
    Center project name: L15250
    Center clone name: 133_A_3
--------------.
* NOTE: This record contains 76 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      680: contig of 680 bp in length
*      681      780: gap of 100 bp
*      781     1490: contig of 710 bp in length
*     1491     1590: gap of 100 bp
*     1591     2312: contig of 722 bp in length
*     2313     2412: gap of 100 bp
*     2413     3099: contig of 687 bp in length
*     3100     3199: gap of 100 bp
*     3200     3899: contig of 700 bp in length
*     3900     3999: gap of 100 bp
*     4000     4692: contig of 693 bp in length
*     4693     4792: gap of 100 bp
*     4793     5495: contig of 703 bp in length
*     5496     5595: gap of 100 bp
*     5596     6310: contig of 715 bp in length
*     6311     6410: gap of 100 bp
*     6411     7105: contig of 695 bp in length
*     7106     7205: gap of 100 bp
*     7206     7933: contig of 728 bp in length
*     7934     8033: gap of 100 bp
*     8034     8719: contig of 686 bp in length
*     8720     8819: gap of 100 bp
*     8820     9539: contig of 720 bp in length
*     9540     9639: gap of 100 bp
*     9640    10367: contig of 728 bp in length
*    10368    10467: gap of 100 bp
*    10468    11158: contig of 691 bp in length
*    11159    11258: gap of 100 bp
*    11259    11965: contig of 707 bp in length
*    11966    12065: gap of 100 bp
*    12066    12758: contig of 693 bp in length
*    12759    12858: gap of 100 bp
*    12859    13557: contig of 699 bp in length
*    13558    13657: gap of 100 bp
*    13658    14362: contig of 705 bp in length
*    14363    14462: gap of 100 bp
*    14463    15183: contig of 721 bp in length
*    15184    15283: gap of 100 bp
*    15284    16051: contig of 768 bp in length
*    16052    16151: gap of 100 bp
*    16152    16861: contig of 710 bp in length
*    16862    16961: gap of 100 bp
*    16962    17661: contig of 700 bp in length
*    17662    17761: gap of 100 bp
*    17762    18448: contig of 687 bp in length
*    18449    18548: gap of 100 bp
*    18549    19259: contig of 711 bp in length
*    19260    19359: gap of 100 bp
*    19360    20087: contig of 728 bp in length
*    20088    20187: gap of 100 bp
*    20188    20876: contig of 689 bp in length
*    20877    20976: gap of 100 bp
*    20977    21670: contig of 694 bp in length
*    21671    21770: gap of 100 bp
*    21771    22420: contig of 650 bp in length
*    22421    22520: gap of 100 bp
*    22521    23227: contig of 707 bp in length
*    23228    23327: gap of 100 bp
*    23328    24093: contig of 766 bp in length
*    24094    24193: gap of 100 bp
*    24194    24910: contig of 717 bp in length
*    24911    25010: gap of 100 bp
*    25011    25680: contig of 670 bp in length
*    25681    25780: gap of 100 bp
*    25781    26450: contig of 670 bp in length
*    26451    26550: gap of 100 bp
*    26551    27226: contig of 676 bp in length
*    27227    27326: gap of 100 bp
*    27327    28029: contig of 703 bp in length
*    28030    28129: gap of 100 bp
*    28130    28850: contig of 721 bp in length
*    28851    28950: gap of 100 bp
*    28951    29642: contig of 692 bp in length
*    29643    29742: gap of 100 bp
*    29743    30448: contig of 706 bp in length
*    30449    30548: gap of 100 bp
*    30549    31252: contig of 704 bp in length
*    31253    31352: gap of 100 bp
*    31353    32039: contig of 687 bp in length
*    32040    32139: gap of 100 bp
*    32140    32849: contig of 710 bp in length
*    32850    32949: gap of 100 bp
*    32950    33659: contig of 710 bp in length
*    33660    33759: gap of 100 bp
*    33760    34476: contig of 717 bp in length
*    34477    34576: gap of 100 bp
*    34577    35266: contig of 690 bp in length
*    35267    35366: gap of 100 bp
*    35367    36074: contig of 708 bp in length
*    36075    36174: gap of 100 bp
*    36175    36854: contig of 680 bp in length
*    36855    36954: gap of 100 bp
*    36955    37650: contig of 696 bp in length
*    37651    37750: gap of 100 bp
*    37751    38452: contig of 702 bp in length
*    38453    38552: gap of 100 bp
*    38553    39269: contig of 717 bp in length
*    39270    39369: gap of 100 bp
*    39370    40046: contig of 677 bp in length
*    40047    40146: gap of 100 bp
*    40147    40844: contig of 698 bp in length
*    40845    40944: gap of 100 bp
*    40945    41652: contig of 708 bp in length
*    41653    41752: gap of 100 bp
*    41753    42478: contig of 726 bp in length
*    42479    42578: gap of 100 bp
*    42579    43272: contig of 694 bp in length
*    43273    43372: gap of 100 bp
*    43373    44053: contig of 681 bp in length
*    44054    44153: gap of 100 bp
*    44154    44854: contig of 701 bp in length
*    44855    44954: gap of 100 bp
*    44955    45680: contig of 726 bp in length
*    45681    45780: gap of 100 bp
*    45781    46475: contig of 695 bp in length
*    46476    46575: gap of 100 bp
*    46576    47286: contig of 711 bp in length
*    47287    47386: gap of 100 bp
*    47387    48064: contig of 678 bp in length
*    48065    48164: gap of 100 bp
*    48165    48874: contig of 710 bp in length
*    48875    48974: gap of 100 bp
*    48975    49686: contig of 712 bp in length
*    49687    49786: gap of 100 bp
*    49787    50503: contig of 717 bp in length
*    50504    50603: gap of 100 bp
*    50604    51302: contig of 699 bp in length
*    51303    51402: gap of 100 bp
*    51403    52106: contig of 704 bp in length
*    52107    52206: gap of 100 bp
*    52207    52922: contig of 716 bp in length
```

```
            *    52923    53022: gap of 100 bp
            *    53023    53741: contig of 719 bp in length
            *    53742    53841: gap of 100 bp
            *    53842    54545: contig of 704 bp in length
            *    54546    54645: gap of 100 bp
            *    54646    55352: contig of 707 bp in length
            *    55353    55452: gap of 100 bp

  Query Match             65.3%;  Score 19.6;  DB 2;  Length 60978;
  Best Local Similarity   84.6%;  Pred. No. 2.3e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          4 ACTTAATTTGACCGAGTAACATTCGC 29
              |||| ||| |||| ||||| ||||||
Db      47184 ACTTCATTGGACCTAGTAATATTCGC 47159


RESULT 13
AL358234
LOCUS       AL358234              103930 bp    DNA     linear   PRI 22-MAY-2002
DEFINITION  Human DNA sequence from clone RP11-164A7 on chromosome 10, complete
            sequence.
ACCESSION   AL358234
VERSION     AL358234.14  GI:21211697
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 103930)
  AUTHORS   Phillimore,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 25, 2002 this sequence version replaced gi:14269935.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-164A7 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .103930
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-164A7"
                     /clone_lib="RPCI-11.1"
ORIGIN

  Query Match             65.3%;  Score 19.6;  DB 9;  Length 103930;
  Best Local Similarity   84.6%;  Pred. No. 2.3e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATT 26
              || |||||||||| |||| | |||||
Db      92999 CAGACTTAATTTGCCCGAATTACATT 93024


RESULT 14
AC026388_0
WPCOMMENT
  Sequence split into 5 fragments  LOCUS AC026388 Accession AC026388
     Fragment Name      Begin     End
     AC026388_0             1    110000
     AC026388_1        100001    210000
     AC026388_2        200001    310000
     AC026388_3        300001    410000
     AC026388_4        400001    416492
LOCUS       AC026388              416492 bp    DNA     linear   HTG 16-OCT-2001
DEFINITION  Mus musculus chromosome 7 clone RP23-426B15, WORKING DRAFT
            SEQUENCE, 15 unordered pieces.
ACCESSION   AC026388
VERSION     AC026388.17  GI:16118084
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 416492)
  AUTHORS   Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
            Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
            Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
            Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
            Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
            Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
            Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
            Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
            Kovar,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Martin,R.,
            Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
            Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oguh,M., Parish,B.,
            Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
            Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
            Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
            Worley,K. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 416492)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 14, 2001 this sequence version replaced gi:12025600.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: MADM
                Center clone name: RP23-426B15
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 75% of reads
                Chemistry: Dye-terminator Big Dye: 25% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 413035 bases at least Q40
                Consensus quality: 421271 bases at least Q30
                Consensus quality: 425796 bases at least Q20
                Estimated insert size: 414396; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 6.6x in Q20 bases; sum-of-contigs estimation
            --------------
```

```
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   119764: contig of 119764 bp in length
            *  119765   119864: gap of unknown length
            *  119865   177594: contig of 57730 bp in length
            *  177595   177694: gap of unknown length
            *  177695   225261: contig of 47567 bp in length
            *  225262   225361: gap of unknown length
            *  225362   263833: contig of 38472 bp in length
            *  263834   263933: gap of unknown length
            *  263934   298288: contig of 34355 bp in length
            *  298289   298388: gap of unknown length
            *  298389   333082: contig of 34694 bp in length
            *  333083   333182: gap of unknown length
            *  333183   362065: contig of 28883 bp in length
            *  362066   362165: gap of unknown length
            *  362166   380755: contig of 18590 bp in length
            *  380756   380855: gap of unknown length
            *  380856   387924: contig of 7069 bp in length
            *  387925   388024: gap of unknown length
            *  388025   393894: contig of 5870 bp in length
            *  393895   393994: gap of unknown length
            *  393995   400042: contig of 6048 bp in length
            *  400043   400142: gap of unknown length
            *  400143   404818: contig of 4676 bp in length
            *  404819   404918: gap of unknown length
            *  404919   408719: contig of 3801 bp in length
            *  408720   408819: gap of unknown length
            *  408820   413398: contig of 4579 bp in length
            *  413399   413498: gap of unknown length
            *  413499   416492: contig of 2994 bp in length.
FEATURES             Location/Qualifiers
     source          1. .416492
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="7"
                     /clone="RP23-426B15"
ORIGIN

  Query Match             65.3%;  Score 19.6;  DB 2;  Length 110000;
  Best Local Similarity   84.6%;  Pred. No. 2.3e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATT 26
              ||||||| |||||| | ||||| |||
Db      24641 CACACTTCATTTGAACTAGTAAAATT 24666


RESULT 15
AC127467
LOCUS       AC127467               139228 bp    DNA     linear   MAM 03-MAY-2003
DEFINITION  Atelerix albiventris clone LB4-4E15, complete sequence.
ACCESSION   AC127467
VERSION     AC127467.3  GI:30348997
KEYWORDS    HTG.
SOURCE      Atelerix albiventris (middle-African hedgehog)
  ORGANISM  Atelerix albiventris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Insectivora; Erinaceidae; Erinaceinae;
            Atelerix.
REFERENCE   1  (bases 1 to 139228)
  AUTHORS   Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Cariaga,K.,
            Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X.,
            Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
            Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Laric,P., Lee-Lin,S.-Q.,
            Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masiello,C.,
            Maskeri,B., McDowell,J., Paguirigan,C., Pearson,R., Portnoy,M.E.,
            Prasad,A., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K.,
            Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V.,
            Vogt,J.L., Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 139228)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 139228)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   4  (bases 1 to 139228)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAY-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On May 3, 2003 this sequence version replaced gi:22417333.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: due
                 Center clone name: 004E15

            This sequence was finished as follows unless otherwise noted:
            all regions were double-stranded, sequenced with an
            alternate chemistry, or covered by high quality data
            (i.e., phred quality >= 30); an attempt was made to resolve
            all sequencing problems, such as compressions and repeats;
            all regions were covered by at least one plasmid subclone
            or more than one M13 subclone; and the assembly was confirmed
            by restriction digest.

            CLONE LENGTH: This sequence represents the entire insert of
            this clone unless otherwise noted. If there are overlapping
            clones, the overlaps are noted in the beginning and end of
            the Features section.
FEATURES             Location/Qualifiers
     source          1. .139228
                     /organism="Atelerix albiventris"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9368"
                     /clone="LB4-4E15"
                     /clone_lib="LB4"
     misc_feature    34774. .34860
                     /note="single clone coverage"
     misc_feature    139221. .139228
                     /note="single clone coverage"
ORIGIN

  Query Match             65.3%;  Score 19.6;  DB 4;  Length 139228;
  Best Local Similarity   84.6%;  Pred. No. 2.3e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATT 26
              ||||||||| ||| || ||| |||||
Db     107135 CACACTTAAATTGCCCCAGTGACATT 107160


Search completed: July 20, 2004, 00:54:40
Job time : 751.574 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 00:15:28 ; Search time 164.754 Seconds
(without alignments)
773.552 Million cell updates/sec

Title: US-09-831-272-2
Perfect score: 30
Sequence: 1 cacacttaatttgaccgagtaacattcgcc 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	30	100.0	30	3	AAA27965	Aaa27965 Box W1 we
c	2	19.8	66.0	110000	7	ACF67367_07	Continuation (8 of
c	3	19.8	66.0	110000	7	ACF65384_1	Continuation (2 of
c	4	19.6	65.3	556	7	ACD94807	Acd94807 Human col
	5	19.4	64.7	61	2	AAQ98799	Aaq98799 External
c	6	19.2	64.0	18753	4	ABL20682	Abl20682 Drosophil
c	7	19	63.3	14968	3	AAA35221	Aaa35221 Human ade
c	8	19	63.3	14968	3	AAF21343	Aaf21343 Human low
c	9	19	63.3	14968	4	AAD14464	Aad14464 Human IL-
c	10	19	63.3	14968	4	AAD15838	Aad15838 Human int
c	11	19	63.3	14968	7	ABZ97037	Abz97037 Human nuc
c	12	19	63.3	17844	3	AAA35223	Aaa35223 Human ade
c	13	19	63.3	17904	3	AAF21345	Aaf21345 Human low
c	14	19	63.3	17904	7	ABZ97039	Abz97039 Human nuc
c	15	19	63.3	33053	6	ABQ67005	Abq67005 Human ang
c	16	19	63.3	110000	6	ABQ67196_0	Abq67196 Listeria
c	17	19	63.3	110000	6	ABQ69245_24	Continuation (25 o
	18	18.8	62.7	1782	3	AAA96679	Aaa96679 Nucleotid
c	19	18.8	62.7	1988	7	ADA69244	Ada69244 Arabidops
	20	18.8	62.7	2348	3	AAA96678	Aaa96678 Nucleotid
	21	18.6	62.0	868	3	AAA52605	Aaa52605 Eosinophi
	22	18.6	62.0	945	4	AAL26424	Aal26424 Human bre
c	23	18.4	61.3	1506	5	AAS70860	Aas70860 DNA encod
c	24	18.4	61.3	3103	5	AAS85091	Aas85091 DNA encod
c	25	18.4	61.3	7500	6	ABL33115	Abl33115 Human imm
c	26	18.4	61.3	8361	4	ABL05662	Abl05662 Drosophil
	27	18.4	61.3	8869	4	ABL13210	Abl13210 Drosophil
	28	18.4	61.3	65608	6	ABL62910	Abl62910 Breast ca
	29	18.4	61.3	65608	6	ABL64414	Abl64414 Stomach c
	30	18.4	61.3	65608	6	ABL67668	Abl67668 Oesophagu
	31	18.2	60.7	11658	4	ABL17680	Abl17680 Drosophil
c	32	18	60.0	403	6	ABN23600	Abn23600 Human ORF
	33	18	60.0	468	3	AAC02933	Aac02933 Human sec
	34	18	60.0	782	4	AAH08439	Aah08439 Human cDN
	35	18	60.0	4413	4	ABL04492	Abl04492 Drosophil
c	36	18	60.0	5544	6	ABL34621	Abl34621 Human met
c	37	18	60.0	5544	6	ABL70478	Abl70478 Chemicall
c	38	18	60.0	5544	6	AAS61441	Aas61441 Human gen
c	39	18	60.0	5787	7	AAL52041	Aal52041 Broccoli
	40	18	60.0	6820	4	ABL12056	Abl12056 Drosophil
	41	18	60.0	6901	4	ABL12076	Abl12076 Drosophil
	42	18	60.0	237961	6	ABQ80552	Abq80552 Human Can
	43	17.8	59.3	534	4	AAL12387	Aal12387 Human bre
	44	17.8	59.3	797	4	AAL21263	Aal21263 Human bre
c	45	17.8	59.3	10594	3	AAZ49991	Aaz49991 Plasmid p

ALIGNMENTS

RESULT 1
AAA27965
ID AAA27965 standard; DNA; 30 BP.
XX
AC AAA27965;
XX
DT 15-AUG-2000 (first entry)
XX
DE Box W1 weak elicitor-responsive cis-element nucleotide sequence.
XX
KW Box W1; elicitor-responsive cis-element; parsley; PR1 promoter; ss;
KW chimeric promoter; pathogen infection; transgenic plant; resistance;
KW herbicide; local response; genetic engineering; disease resistant crop.
XX
OS Petroselinum crispum.
XX
PN WO200029592-A2.
XX
PD 25-MAY-2000.
XX
PF 12-NOV-1999; 99WO-EP008710.
XX
PR 12-NOV-1998; 98EP-00121160.
PR 27-AUG-1999; 99EP-00116981.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG.
XX
PI Kirsch C, Logemann E, Hahlbrock K, Rushton P, Somssich I;
XX
DR WPI; 2000-387804/33.
XX
PT Chimeric promoters mediating gene expression in plants upon pathogen
PT infection, useful for transgenic plant production comprises at least one
PT cis-acting element to direct elicitor-specific expression.
XX
PS Claim 2; Page 32; 73pp; English.
XX
CC This sequence represents Box W1, a weak elicitor-responsive cis-element
CC from the Parsley PR1 promoter. The present invention relates to chimeric
CC promoters capable of mediating local gene expression in plants upon
CC pathogen infection. The chimeric promoters comprise at least one cis-
CC element (see AAA27964-A27979) capable of directing elicitor-specific
CC expression, and a minimal promoter. The chimeric promoters are useful for
CC producing a transgenic plant which has attained resistance or improved
CC resistance against a pathogen. The cis-acting element, chimeric promoter,
CC recombinant gene encoding the chimeric promoter, vector comprising the

```
CC   chimeric promoter and a compound capable of activating the chimeric
CC   promoter are useful for producing pathogen resistant plants, and for
CC   identifying and/or producing compounds capable of conferring induced
CC   resistance to a pathogen in a plant. A compound which specifically
CC   activates or inhibits genes activated in a plant when attacked by a
CC   pathogen is also useful as a plant protective agent or a herbicide. The
CC   chimeric promoter provides rapid and local response to pathogen attack
CC   but shows negligible activity in uninfected parts of the plants and
CC   therefore is most suitable for the engineering of disease resistant crops
XX
SQ   Sequence 30 BP; 9 A; 9 C; 4 G; 8 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 30;  DB 3;  Length 30;
  Best Local Similarity  100.0%;  Pred. No. 0.00023;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATTCGCC 30
              ||||||||||||||||||||||||||||||
Db          1 CACACTTAATTTGACCGAGTAACATTCGCC 30


RESULT 2
ACF67367_07/c
Continuation (8 of 57) of ACF67367 from base 700001 (Photorhabdus luminescens nucleotide
WP Sequence split into 57 fragments  LOCUS ACF67367 Accession Acf67367
WP     Fragment Name      Begin     End
WP     ACF67367_00            1    110000
WP     ACF67367_01       100001    210000
WP     ACF67367_02       200001    310000
WP     ACF67367_03       300001    410000
WP     ACF67367_04       400001    510000
WP     ACF67367_05       500001    610000
WP     ACF67367_06       600001    710000
WP     ACF67367_07       700001    810000
WP     ACF67367_08       800001    910000
WP     ACF67367_09       900001   1010000
WP     ACF67367_10      1000001   1110000
WP     ACF67367_11      1100001   1210000
WP     ACF67367_12      1200001   1310000
WP     ACF67367_13      1300001   1410000
WP     ACF67367_14      1400001   1510000
WP     ACF67367_15      1500001   1610000
WP     ACF67367_16      1600001   1710000
WP     ACF67367_17      1700001   1810000
WP     ACF67367_18      1800001   1910000
WP     ACF67367_19      1900001   2010000
WP     ACF67367_20      2000001   2110000
WP     ACF67367_21      2100001   2210000
WP     ACF67367_22      2200001   2310000
WP     ACF67367_23      2300001   2410000
WP     ACF67367_24      2400001   2510000
WP     ACF67367_25      2500001   2610000
WP     ACF67367_26      2600001   2710000
WP     ACF67367_27      2700001   2810000
WP     ACF67367_28      2800001   2910000
WP     ACF67367_29      2900001   3010000
WP     ACF67367_30      3000001   3110000
WP     ACF67367_31      3100001   3210000
WP     ACF67367_32      3200001   3310000
WP     ACF67367_33      3300001   3410000
WP     ACF67367_34      3400001   3510000
WP     ACF67367_35      3500001   3610000
WP     ACF67367_36      3600001   3710000
WP     ACF67367_37      3700001   3810000
WP     ACF67367_38      3800001   3910000
WP     ACF67367_39      3900001   4010000
WP     ACF67367_40      4000001   4110000
WP     ACF67367_41      4100001   4210000
WP     ACF67367_42      4200001   4310000
WP     ACF67367_43      4300001   4410000
WP     ACF67367_44      4400001   4510000
WP     ACF67367_45      4500001   4610000
WP     ACF67367_46      4600001   4710000
WP     ACF67367_47      4700001   4810000
WP     ACF67367_48      4800001   4910000
WP     ACF67367_49      4900001   5010000
WP     ACF67367_50      5000001   5110000
WP     ACF67367_51      5100001   5210000
WP     ACF67367_52      5200001   5310000
WP     ACF67367_53      5300001   5410000
WP     ACF67367_54      5400001   5510000
WP     ACF67367_55      5500001   5610000
WP     ACF67367_56      5600001   5648894

  Query Match           66.0%;  Score 19.8;  DB 7;  Length 110000;
  Best Local Similarity  91.3%;  Pred. No. 52;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          5 CTTAATTTGACCGAGTAACATTC 27
              ||||| || ||||||||||||||
Db      10530 CTTAAATTCACCGAGTAACATTC 10508


RESULT 3
ACF65384_1/c
Continuation (2 of 6) of ACF65384 from base 100001 (Photorhabdus luminescens nucleotide
WP Sequence split into 6 fragments  LOCUS ACF65384 Accession Acf65384
WP     Fragment Name      Begin     End
WP     ACF65384_0             1    110000
WP     ACF65384_1        100001    210000
WP     ACF65384_2        200001    310000
WP     ACF65384_3        300001    410000
WP     ACF65384_4        400001    510000
WP     ACF65384_5        500001    530312

  Query Match           66.0%;  Score 19.8;  DB 7;  Length 110000;
  Best Local Similarity  91.3%;  Pred. No. 52;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          5 CTTAATTTGACCGAGTAACATTC 27
              ||||| || ||||||||||||||
Db      10745 CTTAAATTCACCGAGTAACATTC 10723


RESULT 4
ACD94807/c
ID   ACD94807 standard; cDNA; 556 BP.
XX
AC   ACD94807;
XX
DT   23-SEP-2003  (first entry)
XX
DE   Human colon cancer cell expressed cDNA #3219.
XX
KW   Open reading frame detection; genome sequencing; colon cancer;
KW   breast cancer; population genome analysis; genetic shift; cancer;
KW   antibiotic resistance; antibiotic non-tolerance; congenital disease;
KW   agriculture; food crop genome; resistance gene; retrovirus;
KW   influenza virus; eukaryotic pathogen detection; trypanosome; Plasmodium;
KW   gene; ss.
XX
OS   Homo sapiens.
XX
PN   US2002155438-A1.
XX
PD   24-OCT-2002.
XX
PF   27-SEP-1999;   99US-00406117.
XX
PR   20-NOV-1998;   98US-00196716.
XX
PA   (SIMP/) SIMPSON A J G.
PA   (NETO/) NETO E D.
PA   (BREN/) BRENTANI R R.
XX
PI   Simpson AJG,  Neto ED,  Brentani RR;
```

```
XX
DR   WPI; 2003-182626/18.
XX
PT   Determining open reading frames of genome of an organism e.g. a human
PT   suffering from cancer involves use of single oligonucleotide primer at
PT   low stringency for preparing single-stranded cDNA from mRNA of
PT   individual.
XX
PS   Example 9; Page 473; 959pp; English.
XX
CC   The invention describes a method of determining open reading frames in
CC   the genome of organism, comprising contacting mRNA from cell of organism
CC   with a single oligonucleotide primer (I) at low stringency, preparing
CC   single-stranded cDNA by reverse transcribing mRNA with (I), amplifying
CC   cDNA , sequencing the product, and repeating the contacting, preparing
CC   and amplifying steps with different primers and sequencing resulting
CC   nucleic acids. The method is useful for: determining that a known
CC   nucleotide sequence from a genome of an organism corresponds to a
CC   nucleotide sequence of an open reading frame; for preparing a contig,
CC   nucleic acid molecule from a genome of an organism; and for sequencing
CC   all or part of a genome of an organism. mRNA is obtained from mammalian
CC   or human cell which is associated with a pathological condition e.g. a
CC   colon cancer or breast cancer cell. The method is useful for analyses of
CC   populations of subjects and can be used to carry out genetic analyses of
CC   large or small populations. further, it can be used to study living
CC   systems to determine if, e.g. there have been genetic shifts which render
CC   an individual or population more or less likely to be afflicted with
CC   diseases such as cancer, to determine antibiotic resistance or non-
CC   tolerance, and so forth. The method can also be used in the study of
CC   congenital diseases, and the risk of affliction to a foetus, as well as
CC   the study of whether the conditions are likely to be passed to offspring
CC   through ova or sperm. The analyses for pathological conditions can be
CC   carried out in all animals, plants, birds, fish, etc. Using this method,
CC   in the area of agriculture, for example the genomes of food crops can be
CC   studied to determine if resistance genes are present, defects in plant
CC   genomes can also be studied in this way. Similarly, the method permits
CC   determination of the pathogens which integrate into the genome, such as
CC   retroviruses and other integrating viruses such as influenza virus, have
CC   undergone shifts or mutations, which may require different approaches to
CC   therapy. This method is also applied to eukaryotic pathogens, such as
CC   trypanosomes, different types of Plasmodium, etc. The method essentially
CC   eliminates sequencing of non-coding portions. This sequence represents a
CC   polynucleotide isolated from human colon cancer cell cDNA library
XX
SQ   Sequence 556 BP; 206 A; 75 C; 101 G; 174 T; 0 U; 0 Other;

  Query Match           65.3%;  Score 19.6;  DB 7;  Length 556;
  Best Local Similarity   84.6%;  Pred. No. 29;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CACACTTAATTTGACCGAGTAACATT 26
              || |||||||||| |||| | |||||
Db        263 CAGACTTAATTTGCCCGAATTACATT 238


RESULT 5
AAQ98799
ID   AAQ98799 standard; tRNA; 61 BP.
XX
AC   AAQ98799;
XX
DT   29-AUG-1996  (first entry)
XX
DE   External guide sequence EGS-17 for targetting RNase P to CAT mRNA.
XX
KW   external guide sequence; EGS; messenger RNA cleavage;
KW   chloramphenicol acetyltransferase; RNase P; ribonuclease; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   stem_loop       1. .11
FT                   /*tag=  a
FT                   /label= D_stem-loop
FT   stem_loop       12. .28
FT                   /*tag=  b
FT                   /label= anticodon_stem-loop
FT   misc_structure  29. .33
FT                   /*tag=  c
FT                   /label= V_loop
FT   stem_loop       34. .50
FT                   /*tag=  d
FT                   /label= T_stem-loop
FT   misc_structure  51. .61
FT                   /*tag=  e
FT                   /label= amino_acyl_acceptor_stem
XX
PN   WO9524489-A1.
XX
PD   14-SEP-1995.
XX
PF   07-MAR-1995;   95WO-US002816.
XX
PR   07-MAR-1994;   94US-00207547.
PR   18-MAR-1994;   94US-00215082.
XX
PA   (UYYA ) UNIV YALE.
XX
PI   Yuan Y,  Guerrier-Takada C,  Altman S,  Liu F;
XX
DR   WPI; 1995-328280/42.
XX
PT   Targetted ribonuclease P cleavage of RNA using an oligo:nucleotide -
PT   comprising a target recognition sequence and a RNase P binding sequence,
PT   useful for treating cancers and viral and bacterial infections.
XX
PS   Example 5; Page 34; 94pp; English.
XX
CC   Any RNA can be targetted for cleavage by RNase P, using a suitably
CC   designed oligonucleotide as "external guide sequence" (EGS) to form a
CC   hybrid with the target RNA and create a substrate for RNase P cleavage.
CC   The EGSs contain sequences which are complementary to the target RNA and
CC   which form secondary and tertiary structure similar to portions of a tRNA
CC   molecule. An EGS must contain at least 7 nucleotides which base pair with
CC   the target sequence 3' to the intended cleavage site to form a structure
CC   like the amino acyl acceptor stem, nucleotides which base pair to form a
CC   stem and loop similar to the T stem-loop, followed by at least 3
CC   nucleotides that base pair with the target sequence to form a structure
CC   like the dihydroxyuracil stem. In an example, chimeric mRNA-EGS
CC   substrates for guiding RNase P to chloramphenicol acetyltransferase (CAT)
CC   mRNA were identified by an in vitro selection procedure. The present
CC   sequence is the EGS segment of such a chimeric mRNA-EGS(CAT) substrate
XX
SQ   Sequence 61 BP; 17 A; 20 C; 13 G; 0 T; 11 U; 0 Other;

  Query Match           64.7%;  Score 19.4;  DB 2;  Length 61;
  Best Local Similarity   58.6%;  Pred. No. 26;
  Matches   17;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          2 ACACTTAATTTGACCGAGTAACATTCGCC 30
              ||||: ||:::|  |||| ||  ::||||
Db         16 ACACUAAAUUUGCACGAGGAAGGUUCGCC 44


RESULT 6
ABL20682/c
ID   ABL20682 standard; DNA; 18753 BP.
XX
AC   ABL20682;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 13519.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
```

```
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 13519; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 18753 BP; 5401 A; 4009 C; 3969 G; 5374 T; 0 U; 0 Other;

  Query Match             64.0%;  Score 19.2;  DB 4;  Length 18753;
  Best Local Similarity   87.5%;  Pred. No. 76;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 CACTTAATTTGACCGAGTAACATT 26
              ||||||||| ||| ||||| ||||
Db       2327 CACTTAATTGGACTGAGTATCATT 2304


RESULT 7
AAA35221/c
ID   AAA35221 standard; DNA; 14968 BP.
XX
AC   AAA35221;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Human adenosine receptor related polynucleotide 2nd SEQ ID NO:95.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US017712.
XX
PR   03-AUG-1998;   98US-0095212P.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstruction, inflammation, allergies, asthma, hypertension,
PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers.
XX
PS   Disclosure; Page 1260-1264; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an antisense
CC   oligonucleotide (ON) with low adenosine (up to 15%), which targets
CC   nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma,
CC   impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of the
CC   ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185
CC   sequences are also called SEQ ID NO:1 to 185, but the sequences differ
CC   from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to
CC   AAA33992) are specifically claimed ONs from the present invention. N.B.
CC   Sequences given in the disclosure of the present invention do not match
CC   up with their corresponding SEQ ID NO: sequences given in the sequence
CC   listing
XX
SQ   Sequence 14968 BP; 5109 A; 2436 C; 2757 G; 4666 T; 0 U; 0 Other;

  Query Match             63.3%;  Score 19;  DB 3;  Length 14968;
  Best Local Similarity   81.5%;  Pred. No. 91;
  Matches   22;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACACTTAATTTGACCGAGTAACATTCG 28
              |||||||   ||||| |||||||||| |
Db       2190 ACACTTATACTGACCAAGTAACATTGG 2164


RESULT 8
AAF21343/c
ID   AAF21343 standard; DNA; 14968 BP.
XX
AC   AAF21343;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #2910.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
```

```
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US008020.
XX
PR   06-APR-1999;   99US-0127958P.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not trigger
PT   adenosine receptors during metabolism, useful e.g. for treating cancers
PT   and respiratory obstructions.
XX
PS   Disclosure; Page 1345-1348; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or
CC   surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention
XX
SQ   Sequence 14968 BP; 5108 A; 2436 C; 2758 G; 4666 T; 0 U; 0 Other;

  Query Match          63.3%;  Score 19;  DB 3;  Length 14968;
  Best Local Similarity   81.5%;  Pred. No. 91;
  Matches   22;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACACTTAATTTGACCGAGTAACATTCG 28
              ||||||| |||| |||||||||| |
Db       2190 ACACTTATACTGACCAAGTAACATTGG 2164


RESULT 9
AAD14464/c
ID   AAD14464 standard; DNA; 14968 BP.
XX
AC   AAD14464;
XX
DT   15-NOV-2001  (first entry)
XX
DE   Human IL-15 gene allele.
XX
KW   Human; interleukin 15; IL-15; gene therapy; chromosome 4q31; infection;
KW   drug screening; anthropological lineage; paternity testing; HIV;
KW   Human Immunodeficiency Virus; forensic application; T-cell leukaemia; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   intron          1. .346
FT                   /*tag= b
FT                   /cons_splice= (5'site:NO, 3'site:NO)
FT                   /number= 1
FT   variation       replace(101, T)
FT                   /*tag= o
FT   variation       replace(191, C)
FT                   /*tag= p
FT   variation       replace(192, C)
FT                   /*tag= q
FT   variation       replace(267, T)
FT                   /*tag= r
FT   variation       replace(333, T)
FT                   /*tag= s
FT   CDS             347. .13604
FT                   /*tag= a
FT                   /product= "Human IL-15 protein"
FT   exon            347. .357
FT                   /*tag= c
FT                   /number= 1
FT   intron          358. .1350
FT                   /*tag= d
FT                   /cons_splice= (5'site:NO, 3'site:NO)
FT                   /number= 2
FT   variation       replace(367, A)
FT                   /*tag= t
FT   variation       replace(431, T)
FT                   /*tag= u
FT   variation       replace(504, C)
FT                   /*tag= v
FT   exon            1351. .1448
FT                   /*tag= e
FT                   /number= 2
FT   intron          1449. .2806
FT                   /*tag= f
FT                   /cons_splice= (5'site:NO, 3'site:NO)
FT                   /number= 3
FT   variation       replace(1536, A)
FT                   /*tag= w
FT   exon            2807. .2892
FT                   /*tag= g
FT                   /number= 3
FT   intron          2893. .8695
FT                   /*tag= h
FT                   /number= 4
FT   exon            8696. .8740
FT                   /*tag= i
FT                   /number= 4
FT   intron          8741. .10602
FT                   /*tag= j
FT                   /number= 5
FT   variation       replace(8866, A)
FT                   /*tag= x
FT   variation       replace(10504, G)
FT                   /*tag= y
FT   exon            10603. .10739
FT                   /*tag= k
FT                   /number= 5
FT   intron          10740. .13492
FT                   /*tag= l
FT                   /number= 6
FT                   /cons_splice= (5'site:NO, 3'site:NO)
FT   exon            13493. .13604
FT                   /*tag= m
FT                   /number= 6
```

```
FT   intron          13605. .14968
FT                   /*tag=  n
FT                   /number= 7
FT                   /cons_splice= (5'site:NO, 3'site:NO)
XX
PN   WO200158914-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-US004130.
XX
PR   08-FEB-2000; 2000US-0181059P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Anastasio AE,  Chew A,  Denton RR,  Nandabalan K,  Stephens JC;
XX
DR   WPI; 2001-522460/57.
DR   P-PSDB; AAE08576.
XX
PT   Novel polynucleotides comprising one of 11, PS1-PS11, single nucleotide
PT   polymorphisms in human interleukin-15 gene, and useful for treating
PT   disorders affected by expression of function of interleukin-15 isogene.
XX
PS   Claim 19; Fig 1; 78pp; English.
XX
CC   The present sequence is human interleukin-15 (IL-15) gene allele located
CC   on chromosome 4q31. The polymorphic variants of IL-15 genes are useful
CC   for studying the expression and function of IL-15 and expressing IL-15
CC   protein for use in useful for screening for candidate drugs to treat
CC   diseases related to IL-15 activity. Genotyping or haplotyping an
CC   individual at the novel IL-15 polymorphic sites are useful for studying
CC   population diversity, anthropological lineage, the significance of
CC   diversity and lineage of the phenotypic level, paternity testing,
CC   forensic applications and for identifying associations between IL-15
CC   genetic variation and a trait such as level of drug response or
CC   susceptibility to disease. Identifying an association between a genotype
CC   or haplotype and a trait, is useful for developing diagnostic tests and
CC   therapeutic treatments for infections, human immunodeficiency virus and T
CC   -cell leukaemia. The identification of an association between a clinical
CC   response and a genotype or haplotype (or haplotype pair) for the IL-15
CC   gene may be the basis for designing a diagnostic method to determine
CC   those individuals who will or will not respond to the treatment, or
CC   alternatively, will respond at a lower level and thus may require more
CC   treatment, i.e. a greater dose of a drug. The genotyping or haplotyping
CC   methods are also useful for developing drugs targetting IL-15. The
CC   genotyping and haplotyping methods are also useful in designing clinical
CC   trials. IL-15 DNA is useful for therapeutic purposes for treating
CC   disorders affected by expression of function of novel IL-15 isogene and
CC   also in gene therapy. Expression of an IL-15 isogene may be turned off by
CC   transforming a targetted organ, tissue or cell population of an
CC   expression vector that expresses high levels of untranslatable mRNA for
CC   the isogene
XX
SQ   Sequence 14968 BP; 5108 A; 2436 C; 2758 G; 4666 T; 0 U; 0 Other;

  Query Match             63.3%;  Score 19;  DB 4;  Length 14968;
  Best Local Similarity   81.5%;  Pred. No. 91;
  Matches   22;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 ACACTTAATTTGACCGAGTAACATTCG 28
              |||||||   |||||  |||||||||  |
Db       2190 ACACTTATACTGACCAAGTAACATTGG 2164


RESULT 10
AAD15838/c
ID   AAD15838 standard; DNA; 14968 BP.
XX
AC   AAD15838;
XX
DT   15-NOV-2001  (first entry)
XX
DE   Human interleukin 15 (IL-15) gene.
XX
KW   Human; interleukin 15; IL-15; gene therapy; chromosome 4q31; infection;
KW   drug screening; anthropological lineage; paternity testing; HIV;
KW   Human Immunodeficiency Virus; forensic application; T-cell leukaemia; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    101
FT                   /*tag=  a
FT                   /note= "This degenerate base represents polymorphic site"
FT   misc_feature    191
FT                   /*tag=  b
FT                   /note= "This degenerate base represents polymorphic site"
FT   misc_feature    192
FT                   /*tag=  c
FT                   /note= "This degenerate base represents polymorphic site"
FT   misc_feature    267
FT                   /*tag=  d
FT                   /note= "This degenerate base represents polymorphic site"
FT   misc_feature    333
FT                   /*tag=  e
FT                   /note= "This degenerate base represents polymorphic site"
FT   misc_feature    367
FT                   /*tag=  f
FT                   /note= "This degenerate base represents polymorphic site"
FT   misc_feature    431
FT                   /*tag=  g
FT                   /note= "This degenerate base represents polymorphic site"
FT   misc_feature    504
FT                   /*tag=  h
FT                   /note= "This degenerate base represents polymorphic site"
FT   misc_feature    1536
FT                   /*tag=  i
FT                   /note= "This degenerate base represents polymorphic site"
FT   misc_feature    8866
FT                   /*tag=  j
FT                   /note= "This degenerate base represents polymorphic site"
FT   misc_feature    10504
FT                   /*tag=  k
FT                   /note= "This degenerate base represents polymorphic site"
XX
PN   WO200158914-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-US004130.
XX
PR   08-FEB-2000; 2000US-0181059P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Anastasio AE,  Chew A,  Denton RR,  Nandabalan K,  Stephens JC;
XX
DR   WPI; 2001-522460/57.
XX
PT   Novel polynucleotides comprising one of 11, PS1-PS11, single nucleotide
PT   polymorphisms in human interleukin-15 gene, and useful for treating
PT   disorders affected by expression of function of interleukin-15 isogene.
XX
PS   Disclosure; Page 73-78; 78pp; English.
XX
CC   The present sequence is human interleukin-15 (IL-15) gene located on
CC   chromosome 4q31. The polymorphic variants of IL-15 genes are useful for
CC   studying the expression and function of IL-15 and expressing IL-15
CC   protein for use in useful for screening for candidate drugs to treat
CC   diseases related to IL-15 activity. Genotyping or haplotyping an
CC   individual at the novel IL-15 polymorphic sites are useful for studying
CC   population diversity, anthropological lineage, the significance of
CC   diversity and lineage of the phenotypic level, paternity testing,
CC   forensic applications and for identifying associations between IL-15
CC   genetic variation and a trait such as level of drug response or
```

CC susceptibility to disease. Identifying an association between a genotype
CC or haplotype and a trait, is useful for developing diagnostic tests and
CC therapeutic treatments for infections, human immunodeficiency virus and T
CC -cell leukaemia. The identification of an association between a clinical
CC response and a genotype or haplotype (or haplotype pair) for the IL-15
CC gene may be the basis for designing a diagnostic method to determine
CC those individuals who will or will not respond to the treatment, or
CC alternatively, will respond at a lower level and thus may require more
CC treatment, i.e. a greater dose of a drug. The genotyping or haplotyping
CC methods are also useful for developing drugs targetting IL-15. The
CC genotyping and haplotyping methods are also useful in designing clinical
CC trials. IL-15 DNA is useful for therapeutic purposes for treating
CC disorders affected by expression of function of novel IL-15 isogene and
CC also in gene therapy. Expression of an IL-15 isogene may be turned off by
CC transforming a targetted organ, tissue or cell population of an
CC expression vector that expresses high levels of untranslatable mRNA for
CC the isogene
XX
SQ Sequence 14968 BP; 5107 A; 2432 C; 2755 G; 4663 T; 0 U; 11 Other;

Query Match 63.3%; Score 19; DB 4; Length 14968;
Best Local Similarity 81.5%; Pred. No. 91;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACACTTAATTTGACCGAGTAACATTCG 28
||||||| ||||| ||||||||| |
Db 2190 ACACTTATACTGACCAAGTAACATTGG 2164

RESULT 11
ABZ97037/c
ID ABZ97037 standard; DNA; 14968 BP.
XX
AC ABZ97037;
XX
DT 17-OCT-2003 (first entry)
XX
DE Human nucleic acid sequence.
XX
KW Human; antisense; lung dysfunction; nasal airway dysfunction;
KW antiinflammatory steroid; ubiquinone; antiinflammatory; antiallergic;
KW antiasthmatic; hypotensive; immunosuppressive; cytostatic; gene therapy;
KW antisense gene therapy; respiratory; lung; adenosine sensitivity;
KW adenosine receptor; bronchodilation; bronchoconstriction; lung allergy;
KW lung inflammation; respiratory disease; ds.
XX
OS Homo sapiens.
XX
PN WO200285308-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US013135.
XX
PR 24-APR-2001; 2001US-0286137P.
XX
PA (EPIG-) EPIGENESIS PHARM INC.
XX
PI Nyce JW, Li Y, Sandrasagra A, Katz E, Pabalan J, Aguilar D;
PI Miller S, Tang L, Shahabuddin S;
XX
DR WPI; 2003-229219/22.
XX
PT Pharmaceutical composition for treating ailments associated with impaired
PT respiration, has oligo(s) antisense to specific gene(s) or its
PT corresponding RNAs, and glucocorticoid or non-glucocorticoid steroid or
PT ubiquinone.
XX
PS Disclosure; SEQ ID NO 12279; 872pp; English.
XX
CC The invention relates to a novel pharmaceutical composition, which has a
CC first active agent comprising an oligonucleotide antisense to the
CC initiation codon, coding region, 5' or 3' end genomic flanking regions,
CC 5' and 3' intron-exon junctions, or regions within 2-10 nucleotides of
CC junctions of genes encoding a polypeptide associated with lung and/or
CC nasal airway dysfunction and a second active agent comprising an
CC antiinflammatory steroid and ubiquinone. A composition of the invention
CC has antiinflammatory, antiallergic, antiasthmatic, hypotensive,
CC immunosuppressive, and cytostatic activity. The composition may have a
CC use in antisense gene therapy. The composition is useful for treating or
CC preventing a respiratory, lung or malignant disease or condition, also
CC for enhancing the prophylactic or therapeutic respiratory effect of an
CC antiinflammatory steroid in a subject, for reducing or depleting levels
CC of, or reducing sensitivity to adenosine, reducing levels of adenosine
CC receptor, producing bronchodilation, increasing levels of ubiquinone or
CC lung surfactant in a subject's tissue, or treating bronchoconstriction,
CC lung inflammation, lung allergies, or a respiratory disease or condition.
CC Note: The sequence data for this patent is not represented in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 14968 BP; 5108 A; 2436 C; 2758 G; 4666 T; 0 U; 0 Other;

Query Match 63.3%; Score 19; DB 7; Length 14968;
Best Local Similarity 81.5%; Pred. No. 91;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACACTTAATTTGACCGAGTAACATTCG 28
||||||| ||||| ||||||||| |
Db 2190 ACACTTATACTGACCAAGTAACATTGG 2164

RESULT 12
AAA35223/c
ID AAA35223 standard; DNA; 17844 BP.
XX
AC AAA35223;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:97.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US017712.
XX
PR 03-AUG-1998; 98US-0095212P.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstruction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers.
XX
PS Disclosure; Page 1264-1268; 1343pp; English.
XX
CC The present invention describes a new composition comprising an antisense
CC oligonucleotide (ON) with low adenosine (up to 15%), which targets

CC nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma,
CC impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of the
CC ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185
CC sequences are also called SEQ ID NO:1 to 185, but the sequences differ
CC from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to
CC AAA33992) are specifically claimed ONs from the present invention. N.B.
CC Sequences given in the disclosure of the present invention do not match
CC up with their corresponding SEQ ID NO: sequences given in the sequence
CC listing
XX
SQ Sequence 17844 BP; 6000 A; 2932 C; 3322 G; 5590 T; 0 U; 0 Other;

Query Match 63.3%; Score 19; DB 3; Length 17844;
Best Local Similarity 81.5%; Pred. No. 93;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACACTTAATTTGACCGAGTAACATTCG 28
||||||| ||||| |||||||||| |
Db 4580 ACACTTATACTGACCAAGTAACATTGG 4554

RESULT 13
AAF21345/c
ID AAF21345 standard; DNA; 17904 BP.
XX
AC AAF21345;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2912.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US008020.
XX
PR 06-APR-1999; 99US-0127958P.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not trigger
PT adenosine receptors during metabolism, useful e.g. for treating cancers
PT and respiratory obstructions.
XX
PS Disclosure; Page 1349-1353; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or
CC surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention
XX
SQ Sequence 17904 BP; 6014 A; 2945 C; 3330 G; 5615 T; 0 U; 0 Other;

Query Match 63.3%; Score 19; DB 3; Length 17904;
Best Local Similarity 81.5%; Pred. No. 93;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACACTTAATTTGACCGAGTAACATTCG 28
||||||| ||||| |||||||||| |
Db 4640 ACACTTATACTGACCAAGTAACATTGG 4614

RESULT 14
ABZ97039/c
ID ABZ97039 standard; DNA; 17904 BP.
XX
AC ABZ97039;
XX
DT 17-OCT-2003 (first entry)
XX
DE Human nucleic acid sequence.
XX
KW Human; antisense; lung dysfunction; nasal airway dysfunction;
KW antiinflammatory steroid; ubiquinone; antiinflammatory; antiallergic;
KW antiasthmatic; hypotensive; immunosuppressive; cytostatic; gene therapy;
KW antisense gene therapy; respiratory; lung; adenosine sensitivity;
KW adenosine receptor; bronchodilation; bronchoconstriction; lung allergy;
KW lung inflammation; respiratory disease; ds.
XX
OS Homo sapiens.
XX
PN WO200285308-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US013135.
XX
PR 24-APR-2001; 2001US-0286137P.

XX
PA (EPIG-) EPIGENESIS PHARM INC.
XX
PI Nyce JW, Li Y, Sandrasagra A, Katz E, Pabalan J, Aguilar D;
PI Miller S, Tang L, Shahabuddin S;
XX
DR WPI; 2003-229219/22.
XX
PT Pharmaceutical composition for treating ailments associated with impaired
PT respiration, has oligo(s) antisense to specific gene(s) or its
PT corresponding RNAs, and glucocorticoid or non-glucocorticoid steroid or
PT ubiquinone.
XX
PS Disclosure; SEQ ID NO 12281; 872pp; English.
XX
CC The invention relates to a novel pharmaceutical composition, which has a
CC first active agent comprising an oligonucleotide antisense to the
CC initiation codon, coding region, 5' or 3' end genomic flanking regions,
CC 5' and 3' intron-exon junctions, or regions within 2-10 nucleotides of
CC junctions of genes encoding a polypeptide associated with lung and/or
CC nasal airway dysfunction and a second active agent comprising an
CC antiinflammatory steroid and ubiquinone. A composition of the invention
CC has antiinflammatory, antiallergic, antiasthmatic, hypotensive,
CC immunosuppressive, and cytostatic activity. The composition may have a
CC use in antisense gene therapy. The composition is useful for treating or
CC preventing a respiratory, lung or malignant disease or condition, also
CC for enhancing the prophylactic or therapeutic respiratory effect of an
CC antiinflammatory steroid in a subject, for reducing or depleting levels
CC of, or reducing sensitivity to adenosine, reducing levels of adenosine
CC receptor, producing bronchodilation, increasing levels of ubiquinone or
CC lung surfactant in a subject's tissue, or treating bronchoconstriction,
CC lung inflammation, lung allergies, or a respiratory disease or condition.
CC Note: The sequence data for this patent is not represented in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 17904 BP; 6014 A; 2945 C; 3330 G; 5615 T; 0 U; 0 Other;

Query Match 63.3%; Score 19; DB 7; Length 17904;
Best Local Similarity 81.5%; Pred. No. 93;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACACTTAATTTGACCGAGTAACATTCG 28
||||||| ||||| |||||||||| |
Db 4640 ACACTTATACTGACCAAGTAACATTGG 4614

RESULT 15
ABQ67005/c
ID ABQ67005 standard; DNA; 33053 BP.
XX
AC ABQ67005;
XX
DT 28-AUG-2002 (first entry)
XX
DE Human angiogenesis associated polynucleotide SEQ ID NO 35.
XX
KW Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcerds;
KW macular degeneration; inflammatory bowel disease; Crohn's disease;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiarteriosclerotic; ds.
XX
OS Homo sapiens.
XX
PN WO200246454-A2.
XX
PD 13-JUN-2002.
XX
PF 06-DEC-2001; 2001WO-EP014320.
XX
PR 06-DEC-2000; 2000DE-01061338.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Schächt O;
XX
DR WPI; 2002-500450/53.
XX
PT New nucleic acid fragments from chemically treated angiogenesis-
PT associated genes, useful for determining methylation status, e.g. in
PT diagnosis or treatment of cancer.
XX
PS Claim 1; SEQ ID NO 35; 41pp + Sequence Listing; German.
XX
CC The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also
CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 33053 BP; 9206 A; 283 C; 6714 G; 16850 T; 0 U; 0 Other;

Query Match 63.3%; Score 19; DB 6; Length 33053;
Best Local Similarity 81.5%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 CACTTAATTTGACCGAGTAACATTCGC 29
||||||||| |||| ||||| || |
Db 7969 CACTTAATTTATCCGACTAACAATCAC 7943

Search completed: July 20, 2004, 00:29:10
Job time : 167.754 secs

(This Page Blank Intentionally)